Escola de Camins
Escola Tècnica Superior d'Enginyeria de Camins, Canals i Ports
UPC BARCELONATECH

Resilience and dependencies in the European Transmission Power Grid. A data science and networks approach

TRABAJO FINAL DE MASTER

Trabajo realizado por:

Mar Fernández Marco

Dirigido por:

Carina Gibert Oliveras

Martí Rosas Casals

Master en:

Ciencia y Tecnología de la Sostenibilidad

Barcelona, 14 de July de 2018

Content

Abstract

In this thesis, we aim at assessing the complexity and resilience of the European Transmission Power Grid (ETPG) following a data science and a complex networks approach. We consider open data related to energy policies and infrastructural and economic variables, together with ETPG reliability data (i.e., major failures and blackout data) of most European countries, considering data from a period of 14 years (2002 – 2014). A Data Science approach is used to understand spatio-temporal patterns of failures and blackouts of the ETPG along the different countries and periods. A combination of clustering methods with post-processing interpretation techniques including variables from complex networks topological analysis is applied to understand the factors associated with blackouts and failures in the different regions and temporal periods. An innovative approach in the field of multivariate time series is used to introduce additional covariables into the analysis, by completing the blackout data with additional open data related to energy policies and infrastructural and economic variables. Adding contextual information to time series contribute to a better understanding of the phenomenon. Results offer a novel approach to understand the relation between these variables and to improve our ability to maintain and guarantee the ETPG's resilience, defined by its structural integrity, security of supply and transport efficiency.

Keywords: Pre-processing, clustering, post-processing, multivariate data analysis, spatio-temporal patterns, data science, energy, complex networks

RESUMEN

En esta tesis, nuestro objetivo es evaluar la complejidad y la resiliencia de la red de energía de transmisión europea (ETPG) siguiendo una ciencia de datos y un enfoque de redes complejas. Consideramos datos abiertos relacionados con políticas energéticas y variables de infraestructura y económicas, junto con datos de confiabilidad ETPG (es decir, fallas importantes y datos de bloqueo) de la mayoría de los países europeos, considerando datos de un período de 14 años (2002 - 2014). Un enfoque de Data Science se utiliza para comprender los patrones espacio-temporales de fallas y apagones del ETPG a lo largo de los diferentes países y períodos. Se aplica una combinación de métodos de agrupación con técnicas de interpretación posterior al procesamiento, incluidas variables del análisis topológico de redes complejas, para comprender los factores asociados con apagones y fallas en las diferentes regiones y períodos temporales. Se utiliza un enfoque innovador en el campo de las series de tiempo multivariantes para introducir covariables adicionales en el análisis, completando los datos de

bloqueo con datos abiertos adicionales relacionados con las políticas energéticas y las variables de infraestructura y económicas. Agregar información contextual a series de tiempo contribuye a una mejor comprensión del fenómeno. Los resultados ofrecen un enfoque novedoso para comprender la relación entre estas variables y mejorar nuestra capacidad para mantener y garantizar la resiliencia de ETPG, definida por su integridad estructural, seguridad del suministro y eficiencia del transporte.

Palabras clave: preprocesamiento, agrupamiento, posprocesamiento, análisis de datos multivariantes, patrones espacio-temporales, ciencia de datos, energía, redes complejas

RESUM

En aquesta tesi, es pretén avaluar la complexitat i la resiliència de la xarxa europea de transmissió de potència (ETPG) seguint una aproximació de xarxes de dades i xarxes complexes. Considerem dades obertes relacionades amb polítiques energètiques i variables infraestructurals i econòmiques, juntament amb dades de fiabilitat ETPG (és a dir, fallades importants i dades d'apagada) de la majoria de països europeus, tenint en compte les dades d'un període de 14 anys (2002 - 2014). Un enfocament de Ciència de Dades s'utilitza per comprendre els patrons espaciotemporals d'errors i apagades de l'ETPG al llarg dels diferents països i períodes. S'aplica una combinació de mètodes de clusterització amb tècniques d'interpretació postprocessament, incloses variables d'anàlisi topològica de xarxes complexes per comprendre els factors associats a apagades i fallades en les diferents regions i períodes temporals. Un enfocament innovador en el camp de sèries temporals multivariants s'utilitza per introduir covariables addicionals en l'anàlisi, completant les dades d'apagada amb dades obertes addicionals relacionades amb polítiques energètiques i variables infraestructurals i econòmiques. Afegir informació contextual a sèries temporals contribueix a una millor comprensió del fenomen. Els resultats ofereixen un enfocament innovador per entendre la relació entre aquestes variables i millorar la nostra capacitat de mantenir i garantir la resiliencia de l'ETPG, definida per la seva integritat estructural, la seguretat del subministrament i l'eficiència del transport.

Paraules clau: preprocessament, agrupament, postprocessat, anàlisi de dades multivariants, patrons espaciotemporals, ciències de dades, energia, xarxes complexes

1 Introduction

The term Technological networks (TN) is used to refer to human made networks built with, both, the generic objective to establish means of communication and information exchange or to distribute resources and commodities of any kind (Newman, 2010). Technological networks are characterized by a huge number of heterogeneous and spatially distributed components, usually connected in a nontrivial way, that tend to display functional patterns that cannot be deduced from the analysis of their individual components. A particularly important subset of TNs includes those dealing with energy sources, including the most essential technological characteristic of our times: the power grid (Pagani & Aiello, 2013).

Provided data is available, Data Science (DS) give tools to study the dynamic and spatial behaviour of this particular type of TN and to obtain useful information for a better understanding and characterization of these TNs, contributing to a proper modelling and improving their management. DS leads to the achievement of useful knowledge from heterogeneous and complex data, which can be used in several fields, helping to analyse, understand and even predict events.

Dealing with complex networks as TNs, implies a high level of complexity and that is the reason why DS is necessary to understand, model and predict the processes involved. The special features of this kind of networks demand a careful approach to improve the analysis to be better known, modelled and consequently better managed or controlled (Karina Gibert, Sànchez-Marrè, & Izquierdo, 2016). This objective to extracting useful, valuable and understandable knowledge from complex data, using DS techniques as a tool, is embedded in what we known as Knowledge Discovery of Data (KDD). KDD approach facilitates the integration of different knowledge sources and fields of expertise and the involvement of end-user criteria and stakeholders' points of view in algorithm design and result interpretation, which bridges the gap between data and real effective decision-making levels (K. Gibert, 2014).

This document presents an original research which combines Data Science and Complex Networks (CN) techniques to exploit available data regarding power grid European TN for a better decision support on management. The main objective of this work is to better understand how reasons involved in blackouts are associated with other variables, so we can better undersnatd why they happen and, hopefully, how to predict them, or even to find means to avoid them, in case it is possible (Martí Rosas-Casals & Solé, 2011; Solé et al, 2008). Modelling is crucial in this sense and data is the key to extract valuable information and connexions between events.

Although several data sources provide very valuable information about power grids, obtaining a dataset suitable for a global analysis, including all available relevant indicators for the entire European territory is not easy. Even if, usually, DS techniques are focused on the Data Mining process, pre-processing and post-processing are important and high-consuming phases of DS (Figure 1). Significant pre-processing issues arise before getting data in the proper structure and formats. It is well-known that pre-processing is a critical step of all data science processes and requires special attention to guarantee a proper transformation of data into added value. Combination of several information sources, non-normalised measures, similar data recorded into slightly different variables, important presence of missing data or hundreds of data bases from which the raw data can be extracted, are just some of the problems that scientists face up when dealing with DS in TNs environment. Most of these problems are also present here and they have been tackled in this work.

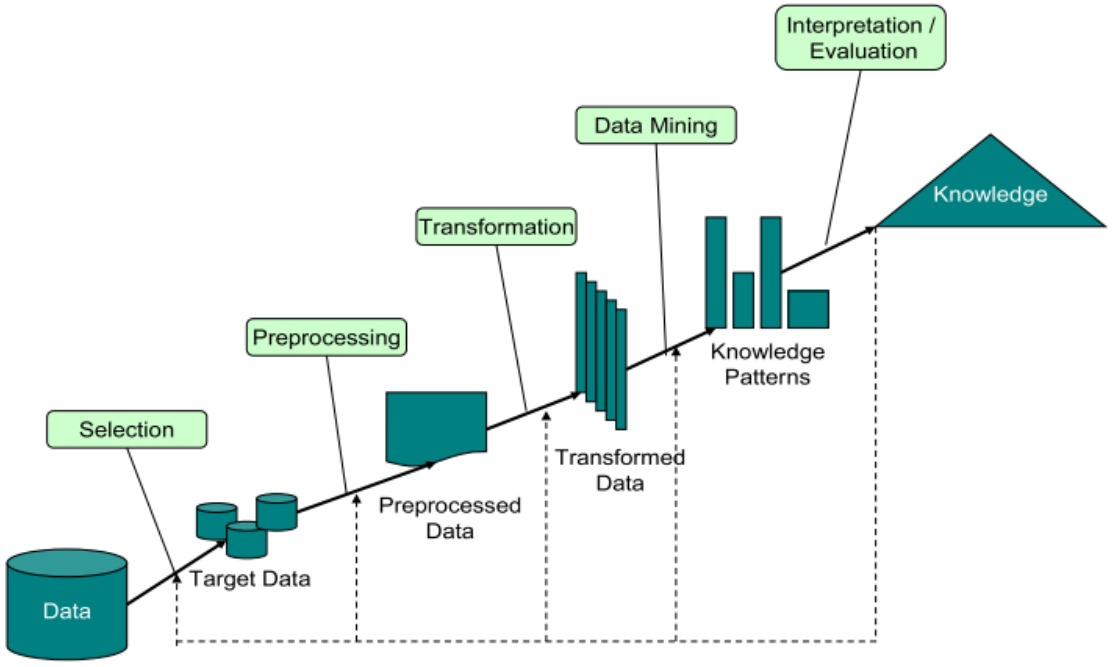


Figure 1. Different steps of the Data Science process, starting from data compilation up to the final use and application of the developed knowledge. Source:(Fayyad, Piatetsky-Shapiro, & Smyth, 1996)

For this particular research and to obtain a suitable working dataset which allows us to learn about power grid patterns, we follow the pre-processing general methodology proposed in Gibert et al. (Gibert, Sànchez-Marrè, & Izquierdo, 2016) (Figure 2)..

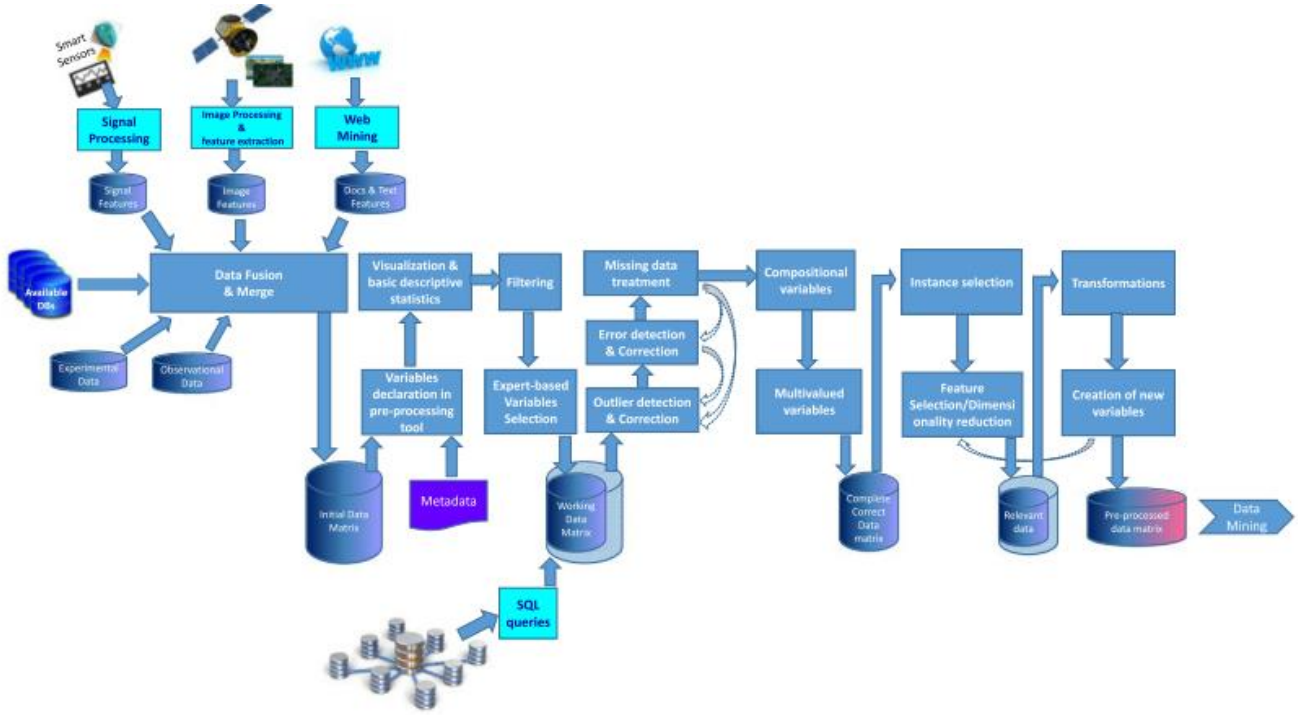


Figure 2. Different steps of the pre-process included in Data Science. Many of the parts have been implemented in this work. The graphic suggest the amount of time that the pre-process step can take in the investigation. Source: (Gibert, Sànchez-Marrè, & Izquierdo, 2016)

The complete study has further horizons. The long-term goal is to contribute to the TN analysis, by means of joining two techniques of complex analysis: Data Science and Complex Networks, which we consider crucial in the last decades and when planning TNs in general, and power grids in particular

Power grids allow the success of advanced economies based on electrical power, but they are also a good example of the limitations imposed by environmental concerns, together with economic and demographic growth (Solé et al., 2008). They are complex networks and they surround us in our daily lives, displaying substantial non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random. They distribute a commodity, electricity, from generation nodes to consumers. Consequently, its spatial distribution in the system is crucial: we talk about national grids, created to satisfy every citizen necessity. As a result, this planar graph tries to achieve the best option to provide electricity to its addressees.

TN structure affects the performance of it, due to the unexpected emergencies that arise from the combination of elements in complex systems. In the special case of the power grid, the present layout of it depends on how it was grown and its evolution (Watts & Strogatz, 1998). Further studies have been developed around the reasons causing and spreading failures in the grid, attending to their distributions, their behaviour as a small-world network, the self-organization dynamics arisen, their

response to random failures and attacks, etc. In spite of the fact that this work does not go so far, we are involving the CN field studying and including some of the typical topological measures of TN considered as graphs. These variables will be later analysed from a statistical and DS point of view, in order to achieve some conclusions that, mixing both disciplines provide with some interesting knowledge.

In this document, we will present, first, the objectives of the work (Section 2). Next step is to explain the methodology followed (Section 3) and results obtained from the study (Section 4). In Section 5 we discuss our results and we finally address our main conclusions achieved in Section 6, together with a proposal of further steps to continue this study in the future.

2 Objectives

The objectives of the present work are diverse as we are dealing with different analysis techniques. Nevertheless, we can identify several general objectives, related to the conclusions that we expect to obtain from the study. The main purpose of the work is to study the European grid of electric transmission in a holistic manner. Even if those grids have been studied from different points of view, here we want to look at them not just as a technical element of a country but as part of its economic and societal structure. The reason why we develop two characterizations of the grid, from a statistical vision and from a complex network one, is trying to understand if some parameters not merely technical are influencing their performance. Moreover, we aim at finding out whether some of the technical parameters might be determinant for the behaviour of the network as an economic and social service. It is interesting to understand if such a network, historically influenced by territorial, political and economic factors, suffers from incidences (i.e. blackouts and other failures) caused by factors that are not completely technical.

Another reason to develop this study on the major events registered in the European transmission grid between 2002 and 2014 is looking into the relationship among different characteristics of the countries and the appearance of failures in the system, with the aim of being able to detect and predict them. This would provide with a more reliable grid and will prevent from technical and economic problems. The characterization and profiling of the countries in the study seeks to find out if there are some variables that define a country, as the total population on it or the GDP per capita that have some effect on the performance and development of the TN. In addition, variables related to the energy policy of each country are studied. We carry out this profiling using DS techniques. This is one of the main objectives of the project, to combine Complex Networks and Data Science techniques in order to obtain a whole vision of the systems and its behaviour. The aim of using both methods searches for integrating the conclusions that we can obtain from the two of them and making the analysis more powerful.

As a conclusion, this study has the main objective to study the failures happened in the European power grids during a recent period of time from both, DS and CN perspective, and joining these two disciplines to characterize these events and help, in next steps, to explain, and maybe even predict, them.

3 Methodology

The steps followed in this study are the following:

1. **Pre-processing:** According to what is described in (Gibert et al, 2016), this first part of the study follows several steps:

1.1. **Database construction:** It requires identification of the information sources containing relevant data in relation with our target problem. Two datasets were built: one providing information for every European country included in the study along the 14 years considered; another one giving information about major events (i.e., blackouts and other failures) and the country where they took place.

1.2. **Data cleaning:** This part of the pre-processing stage is related with formats and data codification encoding of the variables, and has the objective of avoiding miss-matches or correcting errors.

1.3. **Creation of new variables:** New variables are created and added to the datasets by combining those collected from external sources, in order to generate new indicators that fit well in the domain-reasoning schema.

1.4. **Initial descriptive analysis**. It provides a clue on the behaviour of every variable from the two datasets, and helps to design the remaining pre-processing operations required. We will study some statistical parameters of the numerical variables in order to describe the behaviour of each of them:

- **Mean value**: is the sum of the ensemble of values divided by the number of integers on the set.

$$M = \sum_{i=1}^{n} x_i \qquad (1)$$

  Being x_i each of the values taken by the variable and n the total number of values of it.

- **Median value**: value separating the higher half of a dataset from the lower half.

$$Med = \left\{\frac{(n+1)}{2}\right\} th \qquad (2)$$

  Being n the total number of values of the variable and th the element.

- **Standard deviation**: amount of variation or dispersion of the variable.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n-1}} \qquad (3)$$

Being n the total number of values of the variable, x_i each of the values taken by the variable and x the mean M.

Furthermore, qualitative variables are studied by analysing their different modalities, the frequency of appearance of each of these modalities and their proportion of appearance. A list with the sort of the modalities is also represented.

For every numeric variable, a histogram and a boxplot are shown. When possible, regarding to the nature of the variable and its characteristics, we add a time series plot and a plot showing the different values taken by the variable for each year of study. For every categorical variable, we print a barplot and a pieplot. For the variables ENS, TLP, RT and ETI (in the *MEDB*) the histogram of their logarithm is also represented. The reason is that these variables usually take values close to zero what makes difficult to observe dissimilarities andusing the logarithmic transformation helps to their visualization.

1.5. **Missing data imputation:** the Mixed Intelligent-Multivariate Missing Imputation (MIMMI) has been implemented (K. Gibert, 2014). The MIMMI method "*combines expert knowledge and multivariate analysis and is based on clustering procedures. This method does not need to make any assumptions on the distribution models of the variables. This method uses the conditional mean according to the self-underlying structure of the dataset […]. The method does not imply drastic reductions of the variances of estimates after imputation."* The difference between the original method proposed in (K. Gibert, 2014) and the method used here lies on the expert knowledge that we need in order to (a) input the missing values of the first and (b) reduce the set of variables used to calculate the conditional means. In our case, we work with a well-defined data set with many complete variables, that is, variables with all their real values available and intelligent imputation of the first stage is not required. Therefore, a first clustering with a well-defined subset of complete variables is done in order to obtain conditional means. These conditional means per cluster and year, are used to impute remaining variables as MIMMI imputes according to values of similar country.

In addition, K-Nearest Neighbour (k-nn) method and interpolation methods have been used, depending on the type of missings requiring treatment. K-nn algorithm is a non-parametric method used for regression in our case. It inputs the *k* closest training examples in the dataset, obtaining as output the property value for the variable analysed, being this value the average

of the values of its k nearest neighbours. This step ends up with complete datasets, in which, all the values are available and ready to work with.

1.6. **Final basic descriptive analysis**: We repeat the statistical analysis of the datasets, this time, after implementing the missing data treatment, in order to find out any difference in the behaviour of the variables.

2. **Clustering and profiling of the countries**: a clustering process has been developed with the aim of finding patterns of similar countries in certain periods that could be further associated to the occurrences of major events.

The objective of this statistical analysis is to find common behaviour or patterns between countries that help to understand the appearance of blackouts and the characteristics defining them. Therefore, making groups of countries that behave in a similar way is a useful tool that will lead to a more compact and homogeneous definition of the countries as groups. We apply a cluster profiling to the Country dataset.

For the aggregation of the countries in groups, we used an ascending hierarchical method, which considers at first, a number of groups equal to the number of elements that have to be grouped $k0 = n$. At every iteration, the method allows the reduction to k-1 groups, mutually exclusive, considering the union of every possible pair $k(k-1)/2$ that can be formed. The iterative process leads to a complete grouping of the elements in a single class that contains the n original elements. In this thesis, the Ward's method is used, consisting on grouping at every iteration the pair of classes minimizing the inertia between classes lost with the formation of the new class. In each aggregation, the inertia between classes decreases, while the inertia within classes grows.

In order to implement the hierarchical classification by groups, we need to adopt a metric that leads to a robust classification. We chose to use the likelihood coefficient proposed by Gower:

$$S_{ij} = \frac{\sum_{k=1}^{v} S_{ijk}}{\sum_{k=1}^{v} \delta_{jik}} \qquad (4)$$

The likelihood among individuals (the distance itself) is defined as the mean punctuation taken from all the possible comparatives. With n individuals, (25 countries) the matrix $n\ x\ n$ is formed with the likelihood values between the individuals S_{ij}, being δ_{jik} a constant weight assigned to each variable (Url & Society, 1971).The likelihood among individuals (the distance itself) is defined as the mean punctuation taken from all the possible comparatives. With n individuals, (25 countries) the matrix $n\ x\ n$ is formed with the likelihood values between the individuals S_{ij}.

3. **Principal component analysis (PCA)**: this analysis is aimed at providing a visual understanding of the global relationships between variables related to the major events, as well as reducing the dimensionality of our data, also detecting the main variables influencing the nature and the differences between major events in different countries (Chateau & Lebart, 1996). We develop a PCA regarding to different ensembles of variables in order to obtain the maximum information about the relationships among variables.

Principal Components Analysis consists on a rotation that generate a new coordinate system for the original dataset, where new factorial axis are linear transformations of the original variables such that the biggest projected variance is captured in the first axis, the second one in the second axis, and this way progressively. The process starts building the covariance matrix, a symmetric matrix that counts on a complete base of own vectors (Shlens, 2014).

As a result of the coordinate transformation, we obtain a a matrix including the own values of the diagonalization in its, principal diagonal. These values in the diagonal element capture the information preserved by each principal component in terms of variance. . The bigger the variance, more information provides this component. Following this criteria we choose the components containing more information. Each principal component is more or less influenced by each of the analysed numerical variables, and, consequently, provides more or less information regarding to them. PCA has the advantage of maintaining the characteristics or variables from the data set that contribute the most to the variance.

3.1. **Principal Component Analysis for the Major Events database (MEDB)**: PCA applied to the MEDS. It is performed several times using different groups of variables in order to form the principal components.

3.2. **Principal Component Analysis for the Country database (CDB)**: PCA applied to the CDB. It is performed in order to obtain some behaviour patterns in the countries, regarding the blackouts. We analyse the evolution of the countries along time too.

4. **Complex network analysis**. We will perform it in order to obtain a mapping between variables, major events and structural features of the different power grids, considered as graphs. The grid of each of the countries will be studied as a complex network using the elemental parameters of them to characterize the countries.

In order to characterize the network we will use some topological measures, understood as elemental parameters that provide with information relative to the size, the composition, the connectivity and the behaviour of the network as a whole, but also of every vertex and edge.

A network is defined by the number of vertices on it, that we will call n, and the number of edges, namely, m. The European power grid we are analysing will have as vertices the electric substations and as edges the transmission lines. We will consider the following parameters in order to study the topology of the grid:

- **Geodesic path**: this is a path between two vertices such that no shorter path exists. A path is any sequence of vertices such that every consecutive pair of vertices in the sequence is connected by an edge. Those paths have length, being the number of edges traversed along the path.

- **Diameter**: is the length of the longest geodesic path between any pair of vertices in the network for which a path actually exists.

- **Degree of a vertex**: number of edges connected to a node. It is a centrality measure and represents the "importance" in terms of global connectivity of the node:

$$k_i = \sum_{j=1}^{n} A_{ij} \qquad (5)$$

  Being A_{ij} the elements of the adjacency matrix of the network.

- **Closeness centrality of a vertex**: it is the inverse of the mean geodesic distance, which gives the mean distance between two vertices average over all the vertices in the network.

$$c_i = n / \sum_j d_{ij} \qquad (6)$$

  Being n the number of vertices and d_{ij} the distance between two different nodes.

- **Betweenness centrality**: it measures the extent to which a vertex lies on paths between other vertices. It is given by:

$$x_i = \sum_{st} n_{st}^{i} \qquad (7)$$

Being n_{st}^{i} be 1 if vertex i lies on the geodesic path from s to t and 0 if it does not or if there is no such path.

We will also obtain a map for each of the country grids, formed by substations acting as nodes and wires acting as edges. Nevertheless, the study of the network will be basic and we will not focus on the representation or the research in more complex characteristics of them, which may be affecting the resilience or the grouping processes of the network.

Next step is to transform these topological variables to obtain some statistics and create a new dataset, Country Grids dataset, with the topological information of each country power grid.

5. **Principal component analysis regarding the Complex Network characteristics.** We repeat the PCA including the elemental parameters of the network in the study, with the objective of understanding the influence of the characteristics of the network in the country patterns. The new dataset, Country Grids dataset (*GDB*) is added to the *CDB.*
   Due to the lack of information about the national grid of three of the countries in the study (Cyprus, Finland and Lithuania), we implement the MIMMI method again in order to deal with these missing data.
   After implementing the MIMMI method, we obtain a complete dataset that joins the information available in the *CDB* and *GDB.*
   Using this new dataset, we perform the PCA looking for the patterns characterizing the different countries.

We have mainly used two software packages to develop the whole study. *RStudio*[1] and *NodeXL*[2]. Rstudio is an open source data analysis foundation related to the R, which has been used to perform all the statistical analysis. *NodeXL* is an Excel extension that provides with useful tools to study complex networks and that has been used for the last part of the project.

[1] https://www.rstudio.com/

[2] https://nodexl.codeplex.com/

4 Results

4.1 Pre-processing

4.1.1 Data base construction

The power grid analysed here is composed by all the stations and transmission wires present at the 25 countries included in the study. The 25 countries are Austria, Belgium, Croatia, Cyprus, Czech Republic, Finland, France, Germany, Great Britain, Greece, Hungary, Ireland, Italy, Lithuania, Luxembourg, Macedonia, Netherlands, Poland, Portugal, Romania, Serbia, Slovak Republic, Slovenia, Spain and Switzerland. We do analyse the period from year 2002 to year 2014. Some substations work as generation nodes, while others are mere exchange points between different grids.

We counted on a first database providing the information relative to the major events registered at each country during the study period and their technical characteristics, as well as their location and year of occurrence. From this point, we start to accumulate additional information from open data sources, about variables and indicators, which were chosen following expert-based criteria and levels of previous expertise in the field. As data is the key for a better understanding of the phenomenon, the variables analysed here represent different aspects potentially influencing the different major events in terms of investment, technical conditions, socioeconomic and climatic or geographic characteristics, etc. Data was obtained from different sources such as World Bank database (WB), reports from the European Commission (EC) or the European Network of Transmission System Operators for Electricity[3]. In some other cases, data was not available in form of reports, books or databases and a broad research on the Internet was the single option to obtain the information.

Table 1, shows the variables included in this research, the source and some relevant meta-information.

As said before, we will work with two original databases. The first one, namely *Country dataset (CDB),* with data for every country. As the main goal is to identify prototypical scenarios occurring into countries and we do not want to analyse whereas a country keeps stable along time or not, the rows of this matrix describe the socio-economic situation of a country in a certain year, according to the 40 original variables. This matrix contains a total of 325 rows. The second database, namely *Major Events dataset (MEDB)* contains information about the major events (i.e., blackouts or grid failures). It is formed by 879 rows and 47 original variables. Each row describes a blackouts event, by indicating the station, data and country. Columns give information relative to the blackout or the country and the year

[3] ENTSO-E, https://www.entsoe.eu/

of appearance, like the reason of the blackout, and some technical variables as the power lost during the blackout, the time needed to restore the service after it happened (i.e., average interruption time) or the energy not served. In addition, general information about the country in the specific year is included. *MEDB* is formed by merging the information from both *CDB* and the raw database from the ENTSO-E.

4.1.2 Data cleaning

This step consisted on recoding and correcting some variables. For instance, some alphanumerical variables presented typographic inconsistencies, regarding additional "," or blank spaces that created a difference between the same variable when reading them. Some values were in incorrect positions, as a consequence of the manual data collection of some part of the data and it was necessary to place them correctly. This part of the process was time-consuming as a detailed review was needed in order to detect these mistakes and correct them by hand.

4.1.3 New variables construction

As a consequence of working with two datasets, which provide fundamentally similar information but formally different structure, some new numerical variables must be created. These variables collapse information from one matrix and they are added to the other one with the objective of completing it and making both more useful for the analysis. Specifically, there are two variables that were created with technical information taken from the *MEDB* and that were added to the *CDB*. These are the following:

- *ENSAverage:* It gives the mean value of the energy lost in blackouts for a certain country in a particular year.
- *ETIAverage:* It contains the mean value (using the total number of years analysed) of the equivalent time of the blackouts taking into account the amount of service during 12 months.

It is necessary to highlight that, these two transformed variables (*ETI* and *ENS*), as well as others related with major events, tend to show heavy-tailed (i.e., Pareto, power, log-normal, etc.) distributions (Carreras, et al. 2002; Dobson et al., 2007; Luo & Rosas-Casals, 2015; M. Rosas-Casals & Solé, 2011). This fact implies distributions that have heavier tails than the normal distribution and calculating an average or mean value is not always the most accurate statistical feature to consider. We use here these values as an initial and quick, but somehow rough, approximation to adding information to some other variables. Minimun and maximum values of the period are also considered for a better synthesis.

There are four other variables that have been artificially generated during the pre-process:

- *MinETI and MaxETI:* They are, respectively, the minimum and maximum equivalent time of interruption (from the 14 years analysed) of the service due to blackouts taking into account the amount of service during 12 months in the country.
- *Boperyear*: It expresses the number of blackouts happening in every country during a year period.
- *Bonorm*: It expresses the same value, but divided by the number of stations of the country of study.

Boperyear appears in *MEDB* and *CDB* while *Bonorm* is only included in *CBD*. The reason is that the normalization of the number of blackouts per year and country as a function of the number of the stations in the country provides with general data that helps to characterise the countries but that does not provide with new information when dealing with specific data about the failures in the system.

For a better comprehension of the groups, two new variables were created to be used as illustrative variables in the analysis (i.e. to enrich interpretation of clusters, but not to build them).

- *NsubPop:* It is the number of electrical substations per millions of inhabitants.
- *BOyearPop*: It expresses the amount of blackouts registered in one country and one year per millions inhabitants.

Finally, we have ended up with 47 variables (Table 1) for *CDB* and 48 for *MEDB*.

Variable	Name Marix	Meaning	Type	Measuring Unit	Measuring Procedure
EG.ELC.ACCS.RU.ZS	AccElecRur	Acceso to electricity, rural(Data from 1990 to 2010)	Quantitative	%	World Bank
EG.ELC.ACCS.ZS	AccElecPop	Access to electricity (Data from 1990 to 2010)	Quantitative	%	World Bank
EG.ELC.FOSL.ZS	ElecProdOilGas(	Electricity production from oil, gas and coal sources	Quantitative	%	World Bank
EG.ELC.HYRO.ZS	ElecProdHyd	Electricity production from hydroelectric sources	Quantitative	%	World Bank
EG.ELC.NUCL.ZS	ElecProdNuc	Electric power transmission and distribution losses	Quantitative	%	World Bank
EG.ELC.RNWX.ZS	ElecProdRen	Electricity production from nuclear sources	Quantitative	%	World Bank
EG.ELC.LOSS.ZS	TransDistrLoss	Electricity production from renewable sources, excluding hydroelectric	Quantitative	%	World Bank
EG.IMP.CONS.ZS	EnerImp	Net Energy Imports	Quantitative	%	World Bank
EG.USE.ELEC.KH.PC	ElecPowCons	Electric power consumption	Quantitative	kWh per capita	World Bank
EN.ATM.CO2E.PC	CO2Emis	CO2 emissions	Quantitative	metric tons per capita	World Bank
EN.ATM.GHGT.ZG	TotGHGemis	Total greenhouse gas emissions change from 1990	Quantitative	%	World Bank
NY.GDP.PCAP.PP.CD	GDPperCap	GDP per capita, PPP (Data from 1990 to 2010)	Quantitative	current internationa	World Bank
SI.POV.GINI	GINIindex	GINI index (World Bank estimate) (Data from 1990 to 2010)	Quantitative	adimension al	World Bank
SP.POP.TOTL	TotPop	Population, total (Data from 1990 to 2010)	Quantitative	inhabitants	World Bank
SP.RUR.TOTL.ZS	PopRur	Rural population (Data from 1990 to 2010)	Quantitative	%	World Bank
SP.URB.TOTL	PopUrb	Urban population (Data from 1990 to 2010)	Quantitative	%	World Bank
TX.VAL.FUEL.ZS.UN	FuelExp	Fuel exports	Quantitative	%	World Bank
IC.ELC.TIME	TimeElec	Time required to get electricity	Quantitative	days	World Bank
EG.EGY.PRIM.PP.KD	EnergyInten	Energy intensity level of primary energy	Quantitative	MJ/$2011 PPP GDP	World Bank
Corruption_Rank	CorrupRk	Rank of the countries in a global corruption index	Qualitative	rank	Transparency Int'l
Democracy_Rank	DemocRk	Classification within global democracy ranking	Qualitative		Democracy Index 2014
Temperatura (Average)	TempAverage	Average temperature of the country	Quantitative	ºC	Weatherbase
Precipitation (Average)	PrepAverage	Average precipitation of the country	Quantitative	mm	Weatherbase
Climate	Climate	Dominating climate conditions in populated areas of the country	Qualitative		Köppen climate classification
Island	Island	Country surrounded completely by water, or countries without connections on land route	Qualitative		GoogleMaps
Electricity Distribution Market	ElecDistr	How many companies are in the distribution market	Qualitative		www.eurolectric.org
Electricity Generation Market	ElecGen	Which kind of companies are compiting in the generation market	Qualitative		Internet Research
Electricity Transmission Market	ElecTrans	Which kind of companies are compiting in the transmission market	Qualitative		https://ec.europa.eu/energy/sites/ener/files/documents/quarterly_report_on_european_electricity_markets_q3_2017_finalcover.pdf
Electricity Comercial Market	ElecCome	Which kind of companies are compiting in the comercial market	Qualitative		
Regulated Price	RerPrice	Country with regulated prices of the electricity for households	Qualitative		
Nuclear	Nuclear	Nuclears Plants	Qualitative		Internet Research
OCDE	OCDE	Organization for Economic Cooperation and Development, its objective is to coordinate the economic policies of its members.	Qualitative		Internet Research
Ratification of Paris Agreement	RatParis	International agreement aimed at reducing the emissions of six greenhouse gases that cause global warming by an approximate percentage of at least 5% by 2020	Qualitative		Status of ratification
Type of Government	Government	Organization model of the constitutional power that a State adopts in function of the relation between different kinds of agents of power	Qualitative		Internet Research
EU	EU	European Union	Qualitative		Internet Research
Interconnected Country	Intercon	Belonging to one of the six regional wholesale electricity markets in Europe	Qualitative		www.ec.europa.eu
Reason	Reason	Reason why the blackout was caused	Qualitative		www.entsoe.eu
Energy not supplied	ENS	Energy not supplied during the blackout	Quantitative	MWh	www.entsoe.eu
Total loss of power	TL	Total loss of power during the blackout	Quantitative	MW	www.entsoe.eu
Restoration time	RT	Time required to restore the service	Quantitative	min	www.entsoe.eu
Equivalent time of interruption	ETI	Equivalent time of the blackout taking into account the amount of service during 12 monts ((year [in min] * energy not supplied) / consumption last 12 months)	Quantitative	min	www.entsoe.eu
Number of subestations	Nsub	Number of subestations registered on the country	Quantitative	count	Previous study
Number of black outs per year	Boperyear	Number of blackouts registered in a country during a year period	Quantitative	count	Previous study
Number of black outs per year normalized	Bonorm	Number of blackouts registered in a country during a year period normalized in function of the number of subestations in the country	Quantitative	Ratio	Previous study
Number of subestations per million of people	NsubPop	Number of subestations registered on the country normalised for every 1 million of people	Quantitative	count/millio n people	Previous study
Number of black outs per year and per million of people	BOyearPop	Number of black outs registered in a country during a year period normalised for every 1 million of people	Quantitative	Count/millio n people	Previous study
Maximum equivalent time of interruption	MaxETI	Maximum equivalent time of the black out taking into account the amount of service during 12 monts	Quantitative	min	Previous study
Minimum equivalent time of interruption	MinETI	Minimum equivalent time of the black out taking into account the amount of service during 12 monts	Quantitative	min	Previous study

Table 1.Compilation of all variables included in the CDB and MEDB. The table shows the name of the variable and its codification. In addition a brief description of each variable is given, together with its nature, the measure unit and the source.

4.1.4 Initial basic descriptive analysis

In order to better understand the behaviour of the different variables studied, for both datasets, *CDB* and *MEDB*, a basic descriptive analysis is developed. This analysis includes graphical and numerical information. It shows the evolution of the variables along time and the frequency of appearance of each value. For every numeric variable, a histogram and a boxplot are shown. When possible, regarding to the nature of the variable and its characteristics, we add a time series plot and a plot showing the different values taken by the variable for each year of study. For every categorical variable, we print a barplot and a pieplot. For the variables *ENS*, *TLP*, *RT* and *ETI* (in the *MEDB*) the histogram of their logarithm is also represented. The reason is that these variables usually take values close to zero what makes difficult to observe dissimilarities. This problem is solved when using the logarithmic transformation. As a statistical and numerical analysis, for every variable of study we calculate some statistical indicators, as the mean, the median, the typical deviation or the frequency of appearance. Figures 3 to 14are some examples of this basic descriptive analysis.

For the *CDB*, we have created three different groups of variables, regarding the field they are related to: (1) Economic Variables, (2) Energy Variables and (3) Climate Variables. In addition, we included the variables *Country* and *Year*, which act as identification variables. We decided to apply this classification in order to simplify the basic statistical analysis performed.

1. "Economic" variables include those providing information about the social or economic state of the country. Variables such as the GDP per capita, the GINI Index, the amount of population, if the country belongs or not to the European Union (EU), if they have or not ratified the Paris Agreement, etc. are part of this group. An extent of 10 variables forms it.
2. "Energy" variables account by the energy market structure of the country, its energetic policy and behaviour, as well as giving some information about the physical system and the blackouts registered during the 14-year period. Forming this group we found variables as the CO_2 emissions of the country, the type of distribution or generation market, the percentage of electricity produced using renewable sources of energy or the mean energy not supplied per blackout in a certain year for a certain country. Totally 27 variables are adding energy information to the data set.
3. "Climate" variables is the smaller group. It includes columns that deal with climatic or geographical issues. This way, countries are classified by their climate, if they territory is or not entirely an island, and by their average temperature and precipitations. These four variables give a quick idea of some topographical conditions that may affect the energy behaviour of the country or modify the number or the intensity of the blackouts.

For the *MEDB*, we add one more group of variables. These variables identify the blackouts and provide with the technical information related to them. Variables as the month, year, country, total loss of power during the failure, reason of the blackout, time needed to restore the electrical service, etc. form this new group called Blackouts Characteristics Variables, which contains 10 variables. The other three groups are exactly the same ones as for the first data set, apart from the Energy Variables. In this case, this group does not consider *ENSAverage* and *ETIAverage*.

Figure 3 through 9 show examples of the basic descriptive analysis results obtained from the *CDB*. We choose to show these variables as they represent the three classification groups, the two varieties of variables (quantitative and qualitative) and they perform differently. Some of these variables show a wide spectrum in their values (i.e. Urban Population or Electricity Generation Market).

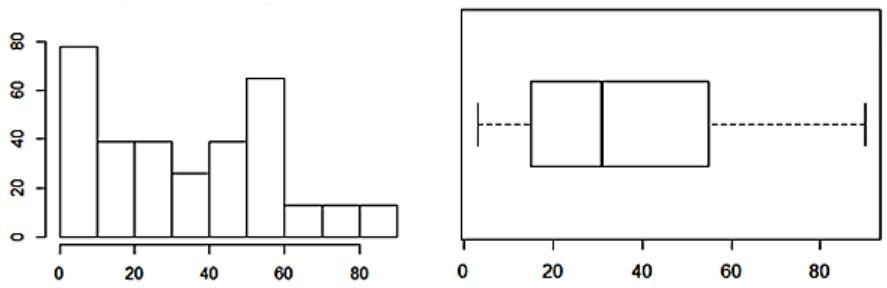


Figure 3. Histogram and Boxplot of the variable "Corruption Ranking". They represent the frequency of appearance, the mean, median, maximum and minimum values of the variable for the CDB.

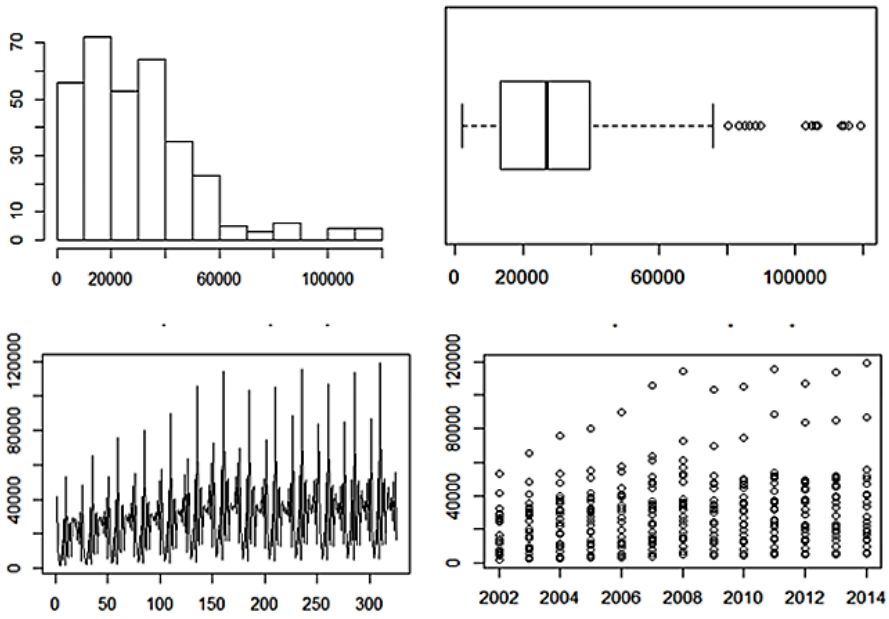


Figure 4. Histogram, Boxplot and Timeplots of the variable "Energy Imports". They represent the frequency of appearance, mean, median, maximum and minimum values of the variables for the CDB as well as its evolution along time.

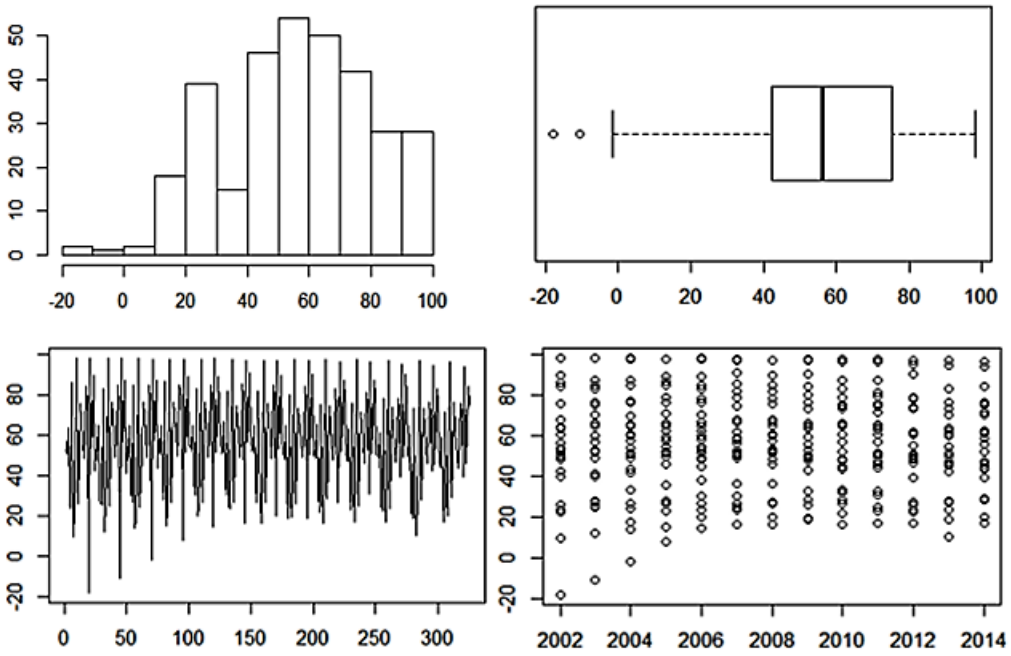


Figure 5. Histogram, Boxplot and Timeplots of the variable "GDP per capita". They represent the frequency of appearance, mean, median, maximum and minimum values of the variables for the CDB as well as its evolution along time.

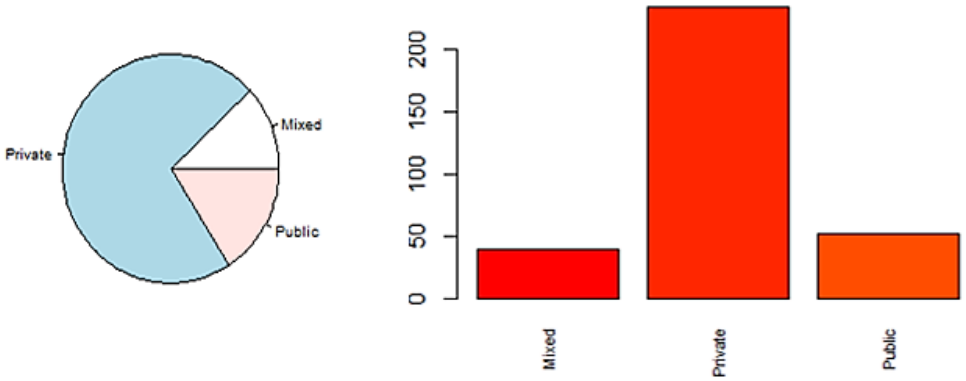


Figure 6. Pieplot and Histogram of the variable "Electricity Generation Market".They represent the frequency of appearance of each of the possible options of the variable for the CDB.

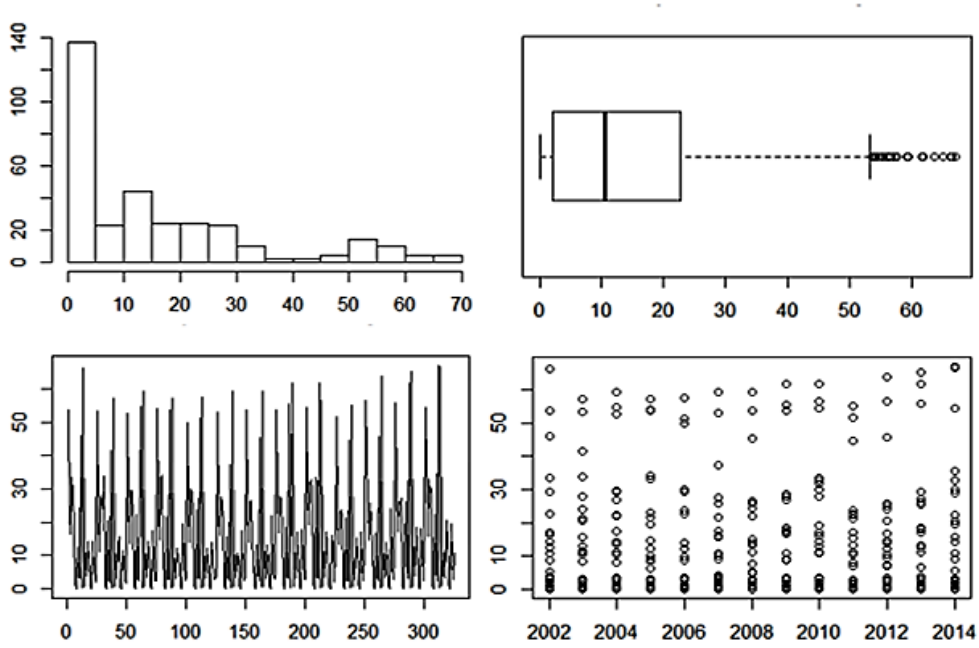


Figure 7. Histogram, Boxplot and Timeplots of the variable "Electricity Production from Hydrological Sources". They represent the frequency of appearance, the mean, median, maximum and minimum values of the variables for the CDB as well as its evolution along time

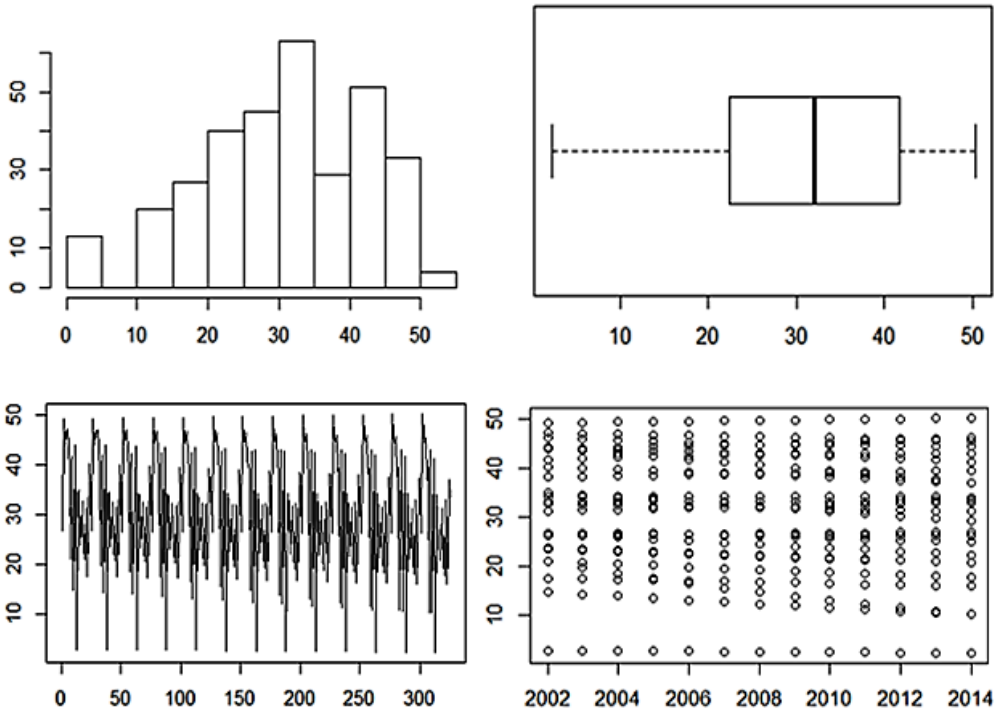


Figure 8.Histogram, Boxplot and Timeplots of the variable "Urban Population". They represent the frequency of appearance, mean, median, maximum and minimum values of the variables for the CDB as well as its evolution along time.

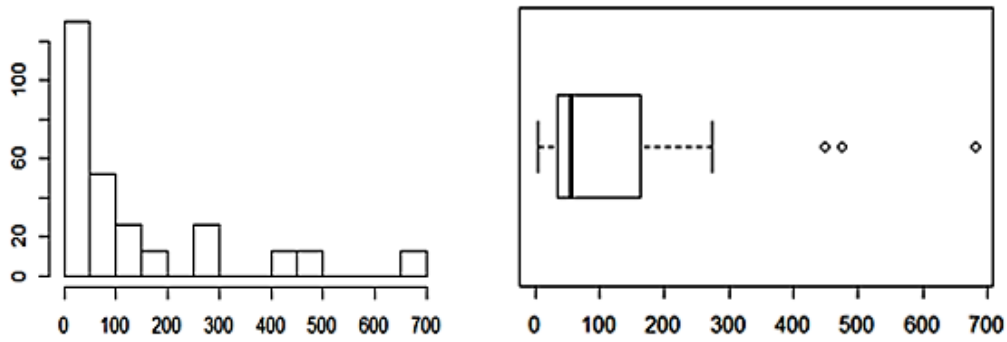


Figure 9.Histogram and Boxplot of the variable "Number of substations". They represent the frequence of appearance, mean, median, maximum and minimum values of the variable for the CDB.

Figure 10 through 14 show examples of the basic descriptive analysis results obtained from the MEDB. These variables are good examples as they represent the four classes and the two types of variables (quantitative and qualitative). In addition they perform differently, helping to understand the general behaviour of the databases. The variable Country appear as a guide of the frequency of appearance of the countries in the dataset. Some of the variables show a wide spectrum in their values (i.e. Reason),

while others are compact in their range of values (i.e. Energy Not supplied during the Blackout or Equivalent Time of Interruption).

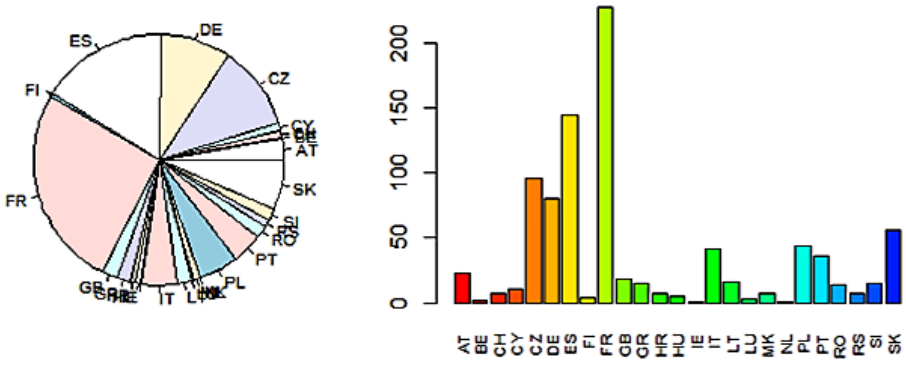


Figure 10.Pieplot and Histogram of the variable "Country". They represent the frequency of appearance of each of the possible options of the variable for the MEDB.

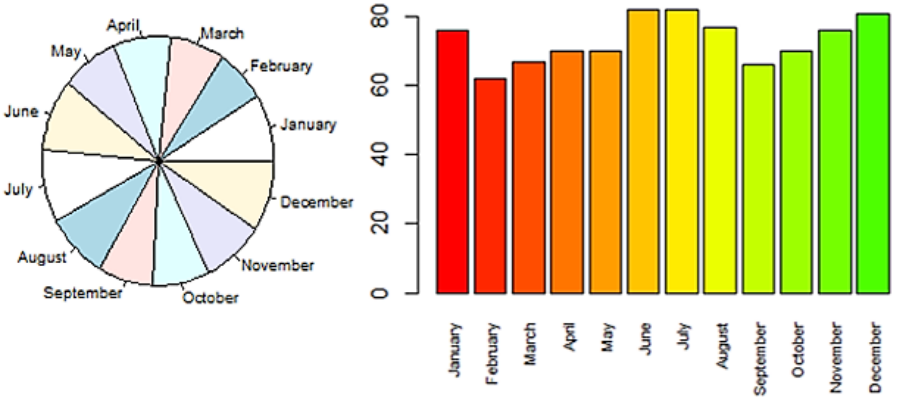


Figure 11.Pieplot and Histogram of the variable "Month". They represent the frequency of appearance of each of the possible options of the variable for the MEDB.

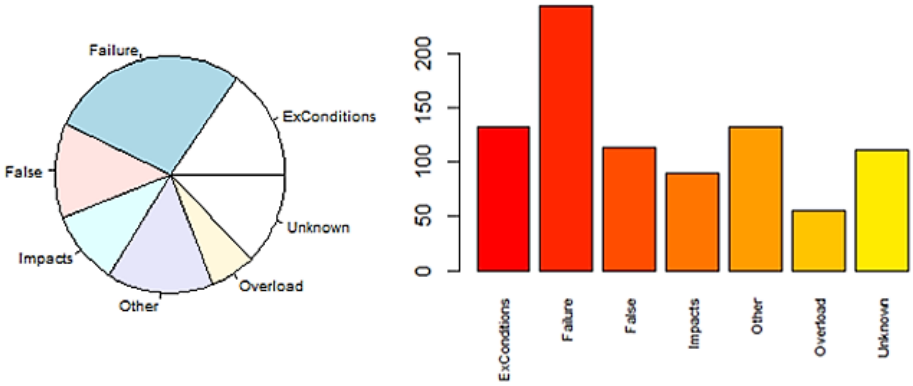


Figure 12. Pieplot and Histogram of the variable "Reason". They represent the frequency of appearance of each of the possible options of the variable for the MEDB.

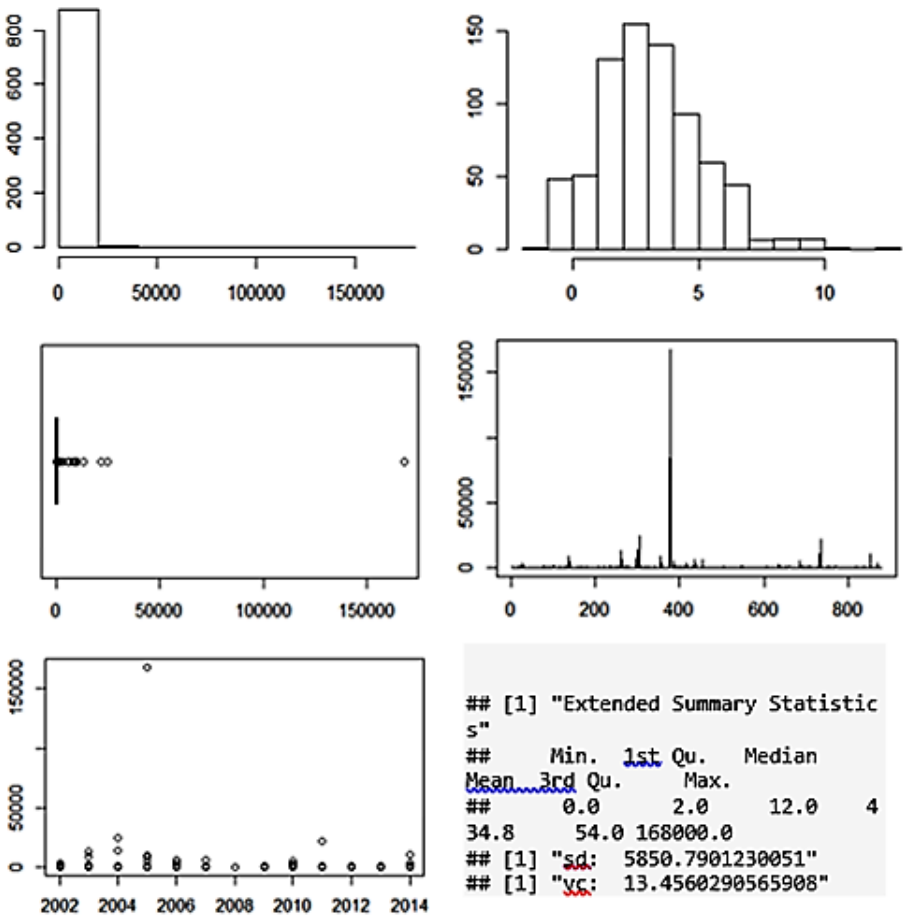


Figure 13. Histograms in linear and logarithmic scales, Boxplot and Timeplots of the variable "Energy Not Supplied during the Blackout". They represent the frequency of appearance, mean, median, maximum and minimum values of the variables for the MEDB as well as its evolution along time. The statistical parameters studied are also represented.

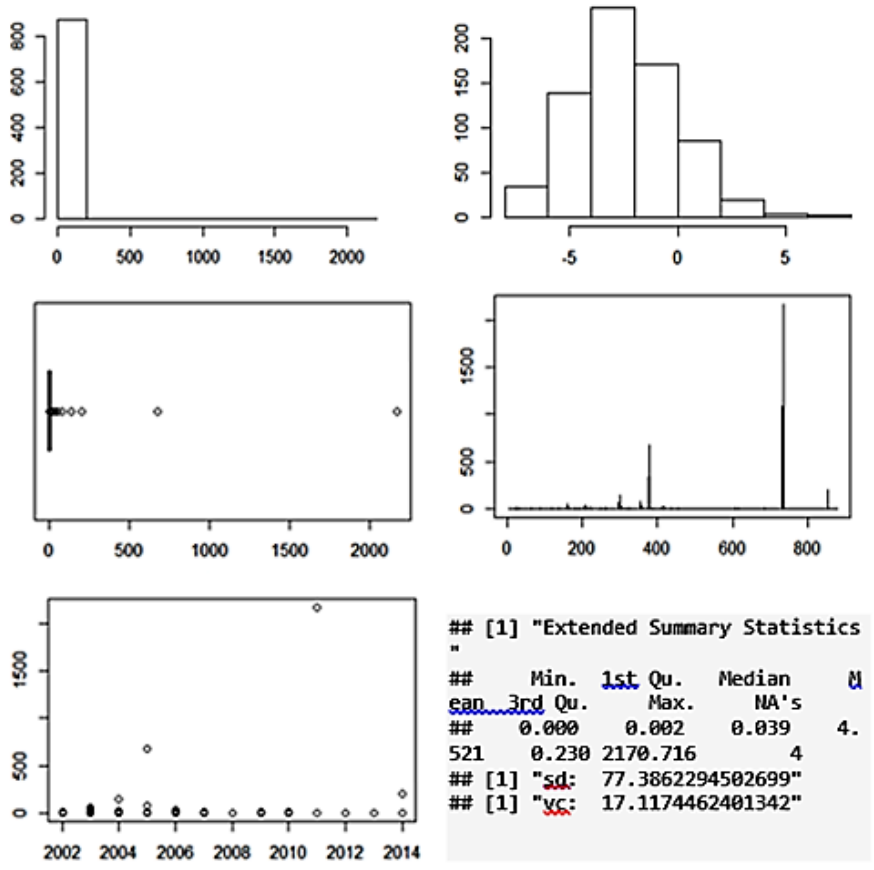


Figure 14.Histograms in linear and logarithmic scales, Boxplot and Timeplots of the variable "Equivalent Time of Interruption". They represent the frequency of appearance, the mean, median, maximum and minimum values of the variables for the MEDB as well as its evolution along time. The statistical parameters studied are also represented.

The complete descriptive analysis for both datasets is available in Appendix A and B.

4.1.5 Missing data imputation

As we mentioned at the beginning of the document, dealing with data related to power grid operation in several countries during a certain period can be difficult if we do not consider all the required data. Luckily, working with European countries and during a recent period makes it easier, and the data matrices are, relatively, quite complete. Nevertheless, there are some values not available. We have to impute this missing in the correct way, if we want to achieve a reliable result of the analysis.

We start the missing data treatment with the *CDB*. The reason is that the *MEDB* is a combination of this matrix and a previous one obtained from the ENTSO-E containing raw data related to the blackouts themselves. As a result, correcting the missing data present at the Country matrix is the first step to

input all the missing values in both databases. Furthermore, we can distinguish between three types of missing data.

1. The first one is missing data associated to the variable itself. There are two variables, *TimeElec* and *TotGHGEmis*, which are not collected every year, so it is not possible to have all the values for each country and year. Those two variables are excluded from the missing treatment and will not be used for creating the clusters neither. We decide to keep these variables as illustrative.
2. The second group includes variables whose missing data depends on the country. That means that, in the case of certain variables, some countries do not provide with any data, independently of the year. This is the behaviour of most of the variables of study, as *CO2Emis*, *FuelExports* or *Nsub*. In order to deal with these not available data we implement the Mixed Intelligent-Multivariate Missing Imputation (MIMMI) with a small modification of the original one presented at Mixed Intelligent-Multivariate Missing Imputation (K. Gibert, 2014). Therefore, a first clustering with a well-defined subset of complete variables provided 6 clusters (Figure 15). Conditional means per cluster and year, are used to impute remaining variables as MIMMI imputes according to values of similar countries.
3. We treat the third group of variables, formed by variables that do not count on values for some specific years, with the k-nn algorithm. This method assigns a value to the elements of the matrix with NA looking for the most similar value or "neighbour" in the matrix and substituting it. The variables that needs this kind of treatment is the *GINIIndex* and, for the *Major Events* dataset the variable *ETI*.
4. Variables punctually missed in some years: In this case, classical interpolation is used.

Once we finished the missing data treatment, we obtained two new datasets with the same number and distribution of rows and columns than the original one, but with no missing values. The exception being variables T*otGHGEmis* and *TimeElec*, which present a different behaviour than the rest.

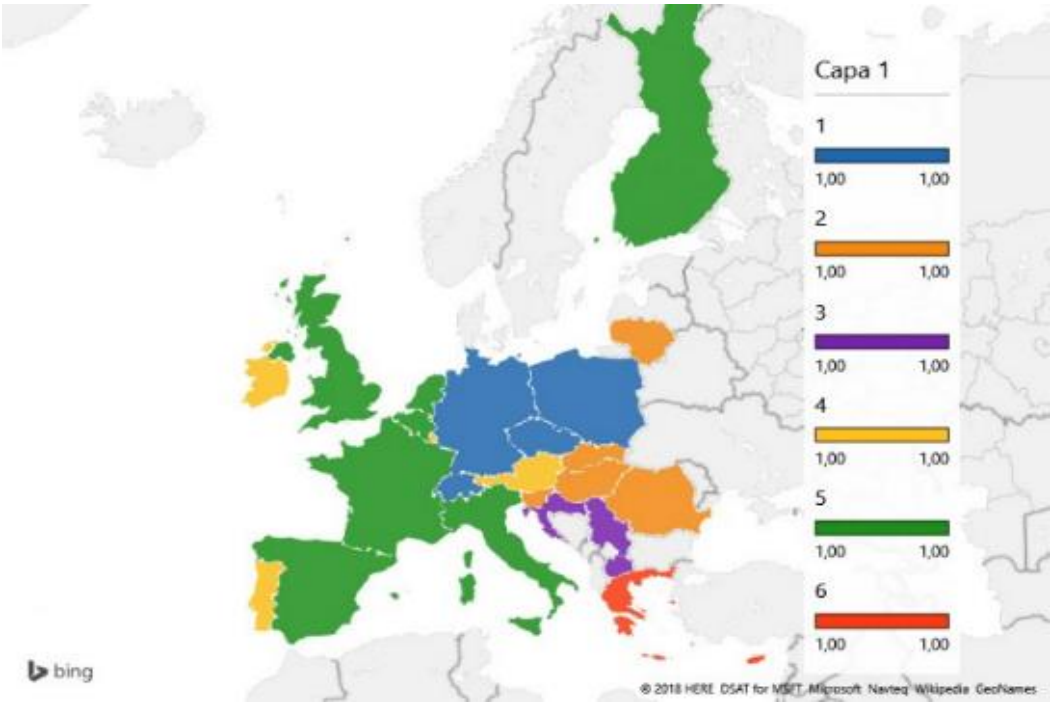


Figure 15. Clusters created with the implementation of the MIMMI method used to impute the missing data. A mean value is calculated for each of the 6 groups and the 14 years of studied. Missing values are substituted by the mean value of their group and year.

4.1.6 Final basic descriptive analysis

The final pre-process step is to repeat the basic descriptive analysis performed before the missing data treatment, this time using the new datasets with all the values available. We do not observe any significate change at the behaviour or tendency of any variable after inputting the missing values. Some examples are shown in Figures 16, 17 and 18.

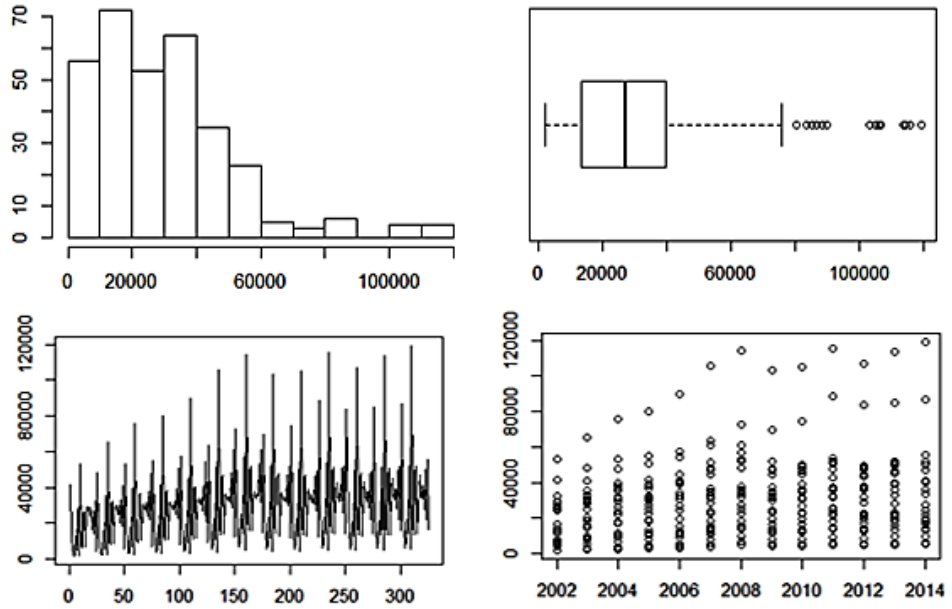


Figure 16. Histogram, Boxplot and Timeplots of the variable "Energy Imports" after missing data imputation. They represent the frequency of appearance, the mean, median, maximum and minimum values of the variables for the CDB as well as its evolution along time.

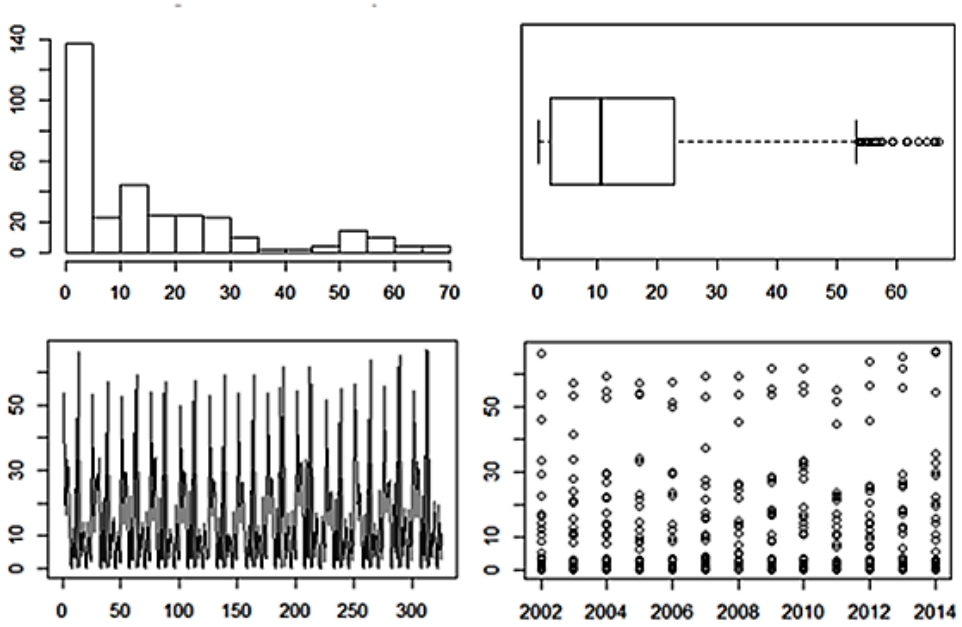


Figure 17. . Histogram, Boxplot and Timeplots of the variable "Electricity Production from Hydrological Sources" after missing data imputation. They represent the frequency of appearance, the mean, median, maximum and minimum values of the variables for the CDB as well as its evolution along time.

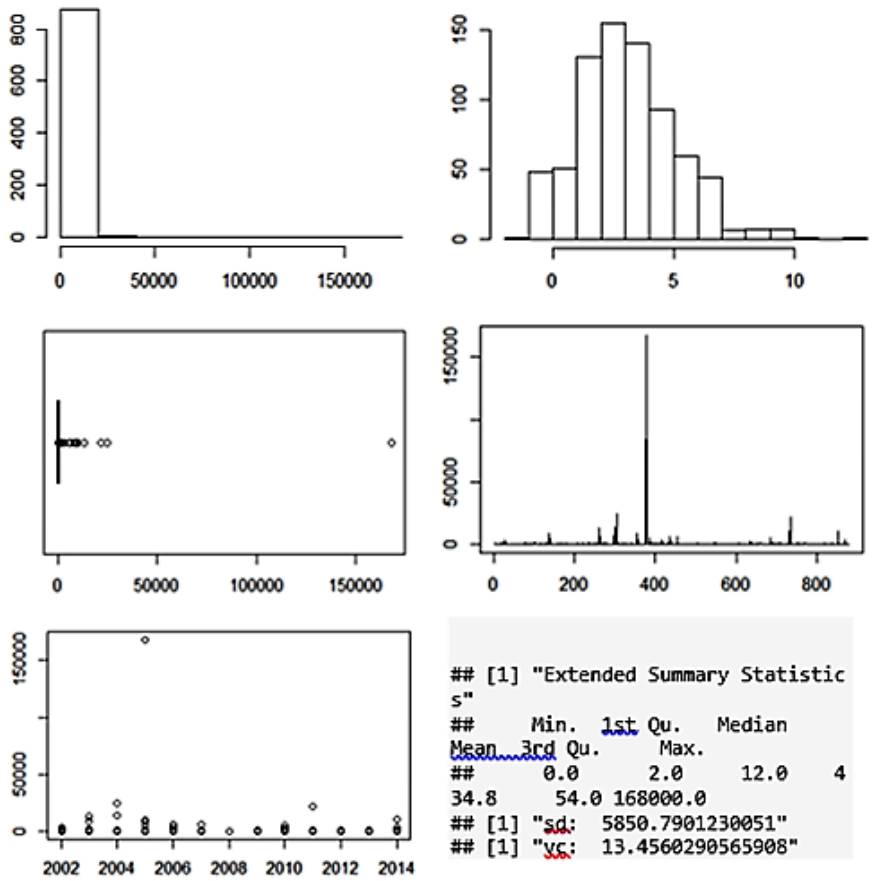


Figure 18. Histograms in linear and logarithmic scales, Boxplot and Timeplots of the variable "Energy Not Supplied" after the missing data imputation. They represent the frequency of appearance, the mean, median, maximum and minimum values of the variables for the MEDB as well as its evolution along time. The statistical parameters studied are also represented.

4.2 Clustering and profiling of the countries

After creating the *CDB* and *MEDB* and developing a pre-process to achieve the proper conditions to work with them, the next step of the analysis is to implement a clustering profiling to the *CDB*. The objective is to create groups of countries with the same characteristics. Analyzing the groups will help to find some clues about the variables affecting the appearance of major events (i.e., major failures and blackouts) in the countries.

All the variables at *CDB* has been considered in the creation of the clusters, with the exception of *TimeElec* and *ToTGHGEmis*. The reason why we put aside these two variables is that, due to their structure, they have a lot of missing values and their use will distort the clustering results.

After implementing the Ward's method to the *CDB*, we obtain a dendogram (Figure 19) that we use to decide the appropriate grouping option.

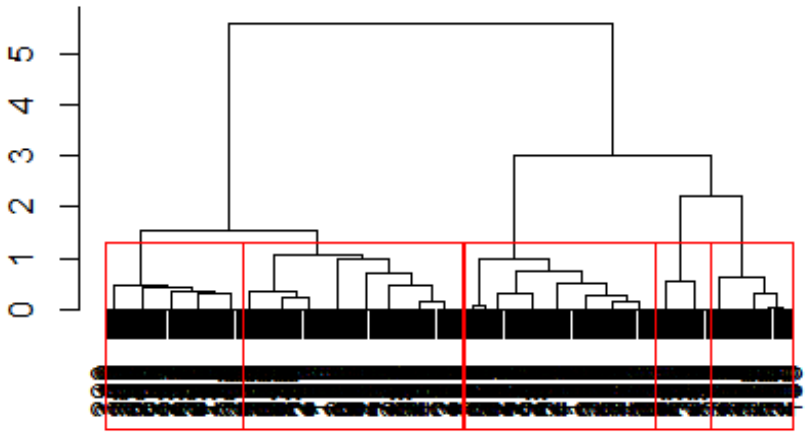


Figure 19. Dendogram used to make decision of how many clusters were the best option to divide, characterize and create profiles of the studied countries.

Using the dendogram as a guidance, we divide the 25 countries into 5 groups. The countries belonging to each group are shown in Table 2 and Figure 20. Every country belongs to the same group during the 14 years analyzed and there is no change of group for any country among the period of study.

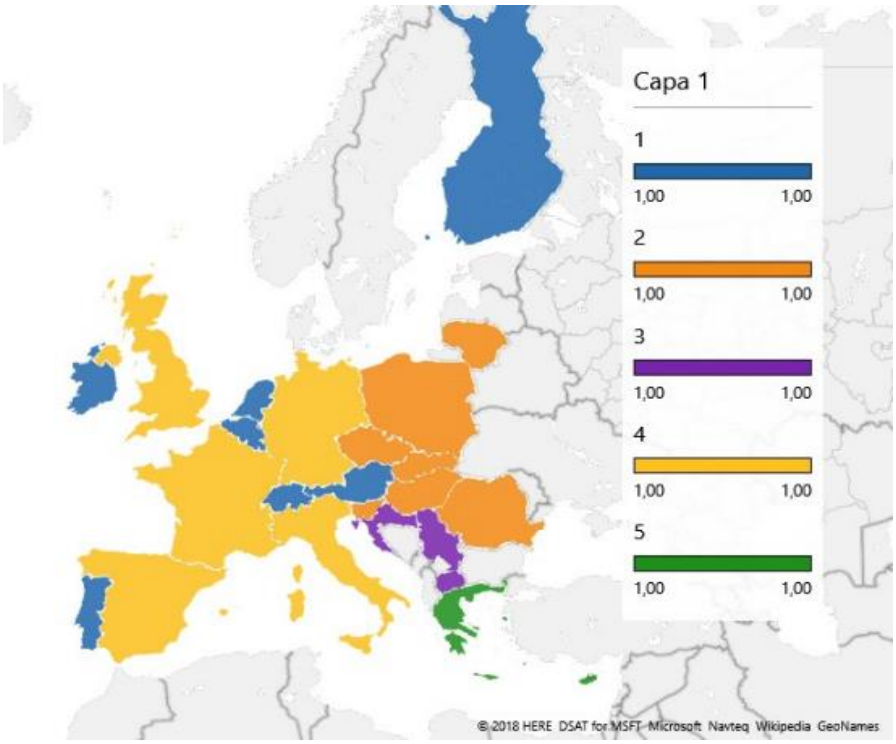


Figure 20. Final cluster division of the 25 countries of study and its composition. Different colours represent the different groups.

1:Northern-Central Europe	2:Eastern Europe	3:Balkan Countries	4:Central and Mediterranean Europe	5:Hellenic Countries
Austria, Belgium, Switzerland, Finland, Ireland, Luxembourg, Netherlands, Portugal.	Czech Republic, Poland, Slovenia, Slovak Republic, Romania, Hungary, Lithuania.	Serbia, Croatia, Macedonia.	France, Germany, Spain, Italy, Great Britain.	Greece, Cyprus.

Table 2. Composition of the five clusters into which countries have been divided in the study.

A basic descriptive analysis is performed for every variable in the *CDB* regarding to its distribution by groups. Here we show the boxplots of the numerical variables (Figure 21, 22 and 23) created for all of them and we provide with some general clues about their behavior attending to the clusters. A deeper and more complete analysis is available in Appendix C.

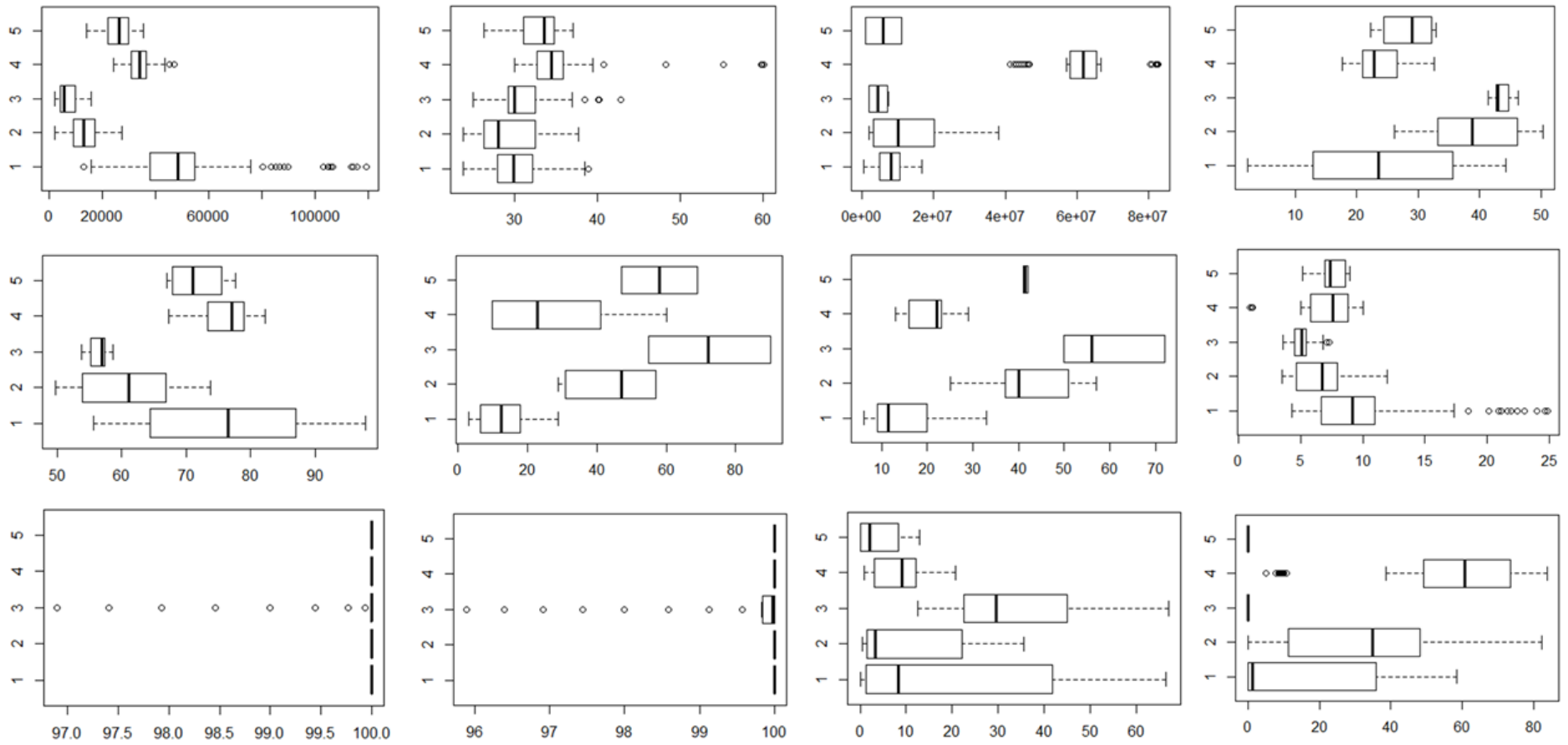


Figure 21. Boxplots showing the mean, median, maximum and minimum values and the distribution regarding the group of belonging of the variables GDPperCap, GINIIndex, TotPop, PopRur, PopUrb, CorrupRk, DemocRk, CO2Emis, AccElecPop, AccElecRur, ElecProdHydr, and ElecProdNuc

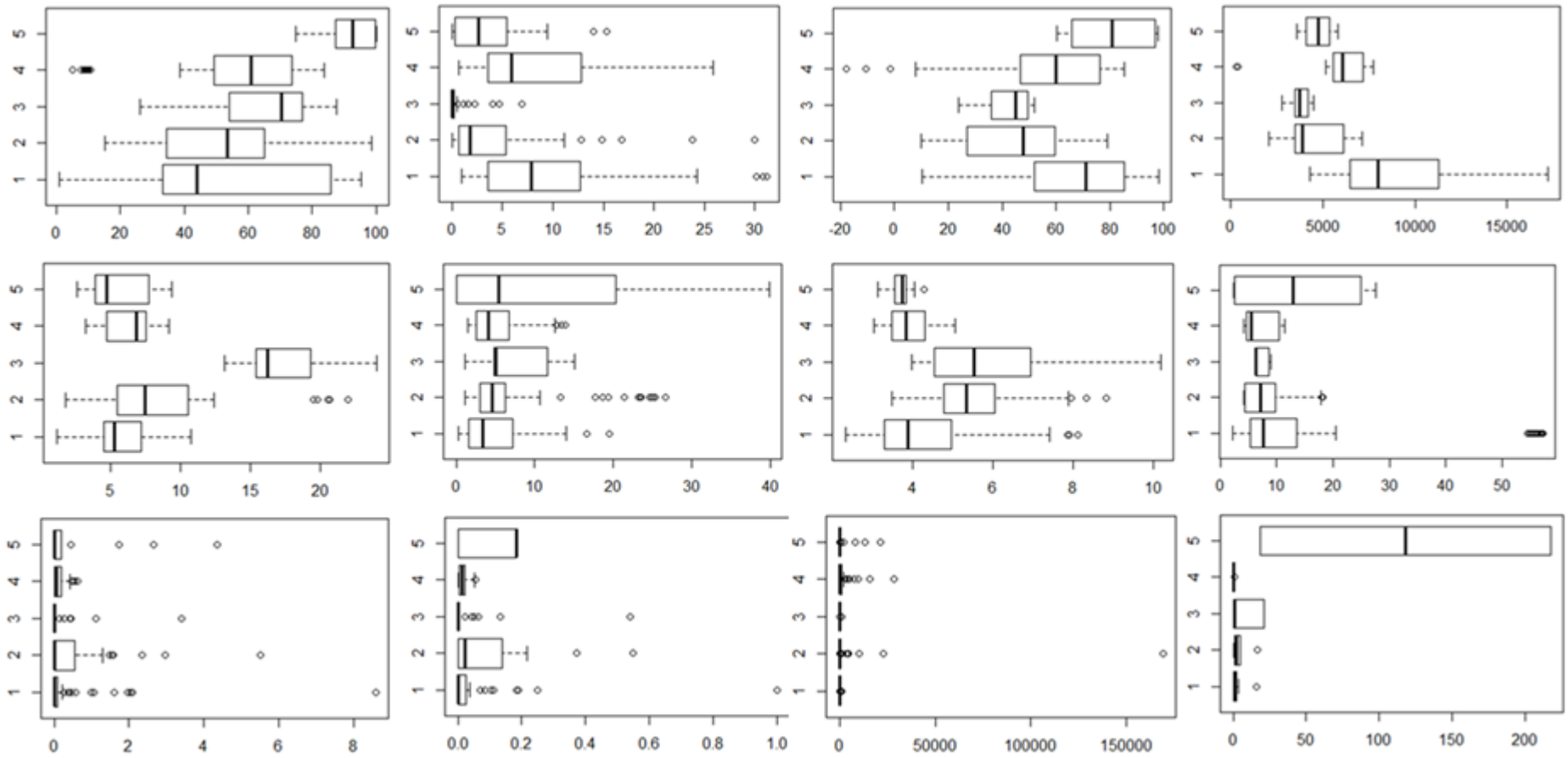


Figure 22. Boxplots showing the mean, median, maximum and minimum values and the distribution regarding the group of belonging of the variables ElecProdOilGasCoal, ElecProdRen, EnerImp, ElecPowCons, TransDistrLoss, FuelExp, EnergyInten, NsubPop, BoperYearPop Nsubnorm, AverageENS and, AverageETI

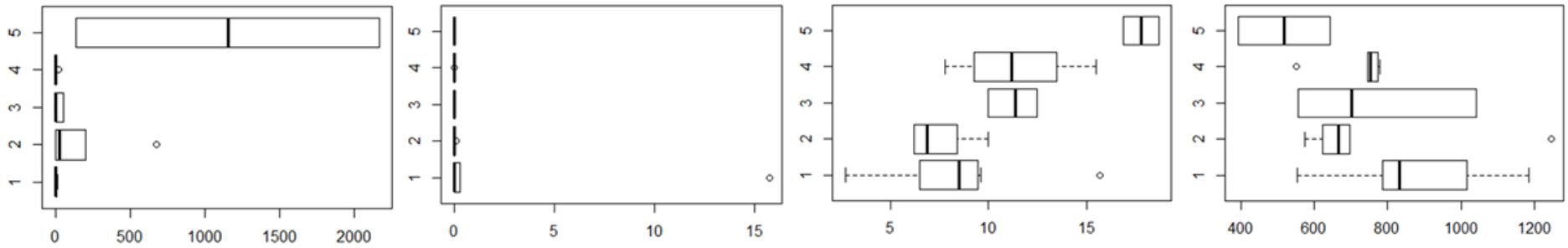


Figure 23. Boxplots showing the mean, median, maximum and minimum values and the distribution regarding the group of belonging of the variables MaxETI, MinETI, AverageTemp and AveragePrep

1. **Northern-Central Europe:** The first group is formed by Austria, Belgium, Switzerland, Finland, Ireland, Luxembourg, Netherlands, and Portugal.
    a. These countries present high levels of GDP per capita. They do not count on big rural populations but the variance of this variable between countries is big. On the contrary, their urban populations are bigger. They also hold the best positions in the democracy and corruption rankings. All the countries belong to the OCDE and to the EU, with the exception of Switzerland.
    b. The CO2 emissions registered during the last years are the most important ones in the whole group of countries, despite of counting on high levels of electricity coming from renewable sources. They form the most electricity-consuming group even if there are big differences between countries. Their energy intensity as a group is smaller than the rest but there are some countries more consuming or less efficient. Their electricity generation systems are entirely private and they all have oligopolistic distribution markets. All of them count on private electricity commercialization and public transmission systems, in this last case, with the exception of Switzerland. Besides, they are all introduced into some wholesale electricity market and the prices of the electricity are liberalized.Regarding the number of substations per millions of people, they count, together with the 5th group, with the bigger results.
    c. Their climates are diverse, although being the oceanic climate the dominant.

    As a conclusion, this group shows a characteristic behaviour for many of the variables although it has a lot of variability among countries. This can be explained by the size of the group, which includes many countries, and by the delocalization of some of them.

2. **Eastern Europe:** Group formed by Czech Republic, Poland, Slovenia, Slovak Republic, Romania, Hungary and Lithuania.
    a. This group includes countries with a low GDP per capita and big levels of rural population. Apart from Romania and Lithuania all countries are in the OCDE and they all belong to the EU. The group holds medium positions at the corruption ranking, as well as at the democracy one.
    b. As a group, they do not produce a lot of electricity from hydrological sources, and they have big variances between countries on the amount of electricity coming from nuclear plants. These countries are not big consumers of electricity and their electricity imports as countries are low. On the contrary, their energy intensities are quite high.

Regarding the distribution market, they all have oligopolistic systems but for the case of Slovenia. Their electricity generation system is form by a mixed matrix of private, public and mixed services. In general, the state owns the transmission grid, with the exception of Czech Republic and Hungary. The commercialization of the electricity is in private hands, excluding Lithuania that has a public managed system. The group is completely interconnected in terms of regional electricity markets.

This group has the biggest amount of blackouts per year when we consider the population of each country. They do not have many substations per population, but on the contrary, they have the biggest amount of blackouts per year per 1000000 of people.

c. All the countries have humid continental climate, and cold temperatures with high variances.

To summarize, the Eastern Group is formed by countries sharing a near localization, climate and political history which is present in their electricity and energy markets.

3. **Balkan Countries:** Serbia, Croatia and Macedonia form the third group.
    a. The countries in this group are all out of the OCDE and their GDP per capita are in the lowest levels among the studied regions. They count on relatively big rural populations and smaller urban ones. They hold the lowest positions in the corruption ranking in Europe.
    b. Their CO_2 emissions are the lowest and there are almost no differences between countries. They do not produce electricity from nuclear sources or renewable ones, but they have a system where hydrological sources are quite relevant. Their electric consume is low and homogeneous for the whole group. They do have the highest levels of energy intensity although there exist dissimilarities between the countries. Their electrical power grids have the biggest losses of the whole set but this behaviour is not compact for the group.

        All the countries have oligopolistic distribution systems and a mixed and private generation market depending on the country. Serbia and Macedonia have a private transmission system, while Croatia manage it publically. Serbia has mixed commercialization and have regulated prices. On the contrary, Croatia and Macedonia count on private commercialization of the electricity and liberalized prices. With the 4^{th} group they have the lowest level of substations and blackouts per year for every 1 million of people. None of them is interconnected with foraging electric systems.

c. About the climatic conditions, they present medium temperatures, a mix of climates and medium level of precipitations.

As a conclusion, these three countries are quite similar between them. This can be caused by their similar climates or their recent history in common. They seem to be at the queue of the European development and present differences with the rest of the countries.

4. **Central and Mediterranean Europe:** In this group, we find France, Germany, Spain, Italy and Great Britain.
    a. Forming this group, we found countries that belong to the OCDE and that have big populations, being more important the urban populations than the rural ones. They count on private electricity generation systems and their GDP per capita is the biggest one in the set. Moreover, they occupy good positions in the corruption and democracy rankings, the best positions after the 1st group.
    b. The electrical system of the group shows the relevance of the nuclear and renewable sources of energy. Nevertheless, the hydrological production is less frequent than in other groups. Their electric consumption in in the medium level and they form the group exporting less amounts of fuel.

       All the countries present oligopolistic distribution markets, private generation, public transmission (apart from Germany) and a private electricity commercialization system. As well, all the countries are part of an international market of electricity, excluding Italy.

       This group have the biggest number of substations in absolute terms but the smallest level if we consider the amount of population. They behave in the same way for the blackouts per year and population.
    c. About climate, they present mixed ones and medium values of the temperatures and precipitations.

This group is formed by the countries heading the economic activity and the European development in many energy policy areas. In addition, they have the oldest and most extended grids in Europe, showing special characteristics as the size or the number of incidences during the year.

5. **Hellenic Countries:** They are Greece and Cyprus.
    a. This group formed by the two Hellenic countries presents a medium-low level of GDP per capita. They are characterise by smaller rural populations and bigger

urban ones. From the rankings, we can conclude that they all are high-corrupted countries with medium levels of democracy development.

b. Greece and Cyprus do not produce electricity from nuclear plants; instead, they have an electric system based on the use of fossil fuels. As a group, they show the biggest energy imports values even if their electric power consumption is not too high. This is the group exporting more fuel to other countries. Regarding the energy intensity, they have relative small values, probably caused by the low levels of electric consumption.

   Both countries have monopolistic distribution systems, public generation markets, and public transmission. About the electricity commercialization, they behave differently, Greece counts on a public managed system while Cyprus has a private one. Greece is in a wholesale electricity market with other regions but Cyprus is not interconnected.

   The average duration time of the black outs in these countries is the biggest in the set, with big dissimilarities to the other groups. Moreover, Greece and Cyprus have the biggest amount (with the 1st group) of electric substations if we consider the population and the highest occurrence of blackouts per year and 1.000.000 of people.

c. They are the warmest and driest countries in the set, sharing a Mediterranean climate.

The cluster profiling has turned out to be a proper characterization tool for the dataset. Through its implementation we have stablish five groups of countries with their own features in terms of economy, climate and policy. The groups are also dissimilar when talking about their Power grids, at least in terms of size. In spite of the fact that all the countries belong to Europe, and most of them to the EU, and they share many characteristics in general terms, we have been able to find some heterogeneities that have been used to group them and make comparisons among clusters.

4.3 Principal components analysis (PCA)

4.3.1 PCA for the Major Events dataset

Next step in the analysis is a Principal Components Analysis. We will start the PCA using the variables in the *MEDB*. Numerical variables are used as active variables, which are those used to compose the Principal Components (PC), while categorical variables will just be used as illustrative. Again, variables *TotGHGEmis* and *TimeElec,* despite of being numerical, are not included in the "actives" group due to the high presence of missing data.

In this first step, the variables used to perform the PC are shown in Table 3:

CO2Emis	ElecProdOilGasCoal	GDPperCap	PopUrb	PrepAverage	RT
AccElecPop	ElecProdRen	GINIindex	EnergyInten	Nsub	ETI
AccElecRur	EnerImp	FuelExp	CorrupRk	Boperyear	
ElecProdHyd	ElecPowCons	TotPop	DemocRk	ENS	
ElecProdNuc	TransDistrLoss	PopRur	TempAverage	TLP	

Table 3. Variables used to perform the first PCA for the MEDB. For this first analysis, all the numerical variables in the MEDB are included actively in the study.

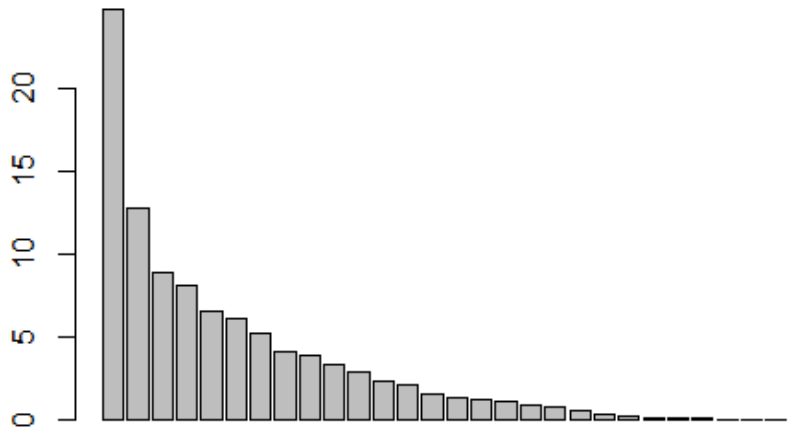


Figure 24 Standard deviation of each of the PCs obtained from the PCA developed using the variables in Table 3

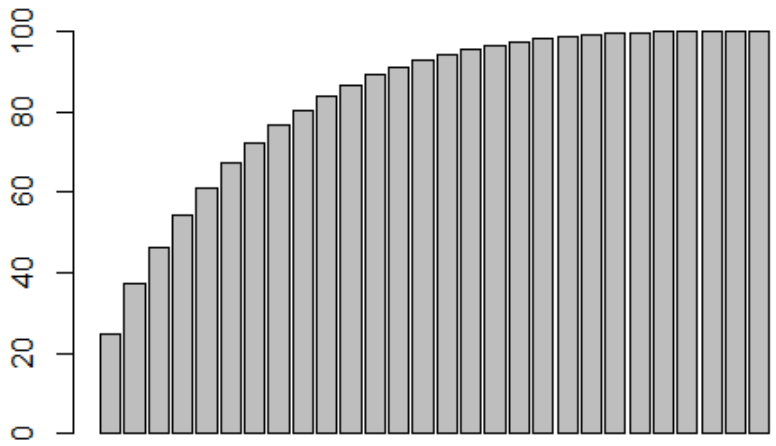


Figure 25 Cumulative percentage of inertia of each of the PCs obtained from the PCA developed using the variables in Table 3

Figure 24 shows the information provided by each PC, in terms of the projected inertia . Figure 25 shows the cumulative percentage of inertia of subspaces of increasing dimensionality, which is a measure of the quantity of information that each subset of first k components keeps. Following the criteria of choosing the minimum subset that keeps at least an 80% of the total information of the original dataset, we decide to work with 9 PCs. Working with the number of components that shows at least the 80% of the available information in the dataset is a general procedure when performing PCA in order to achieve reliable results. In this case, the first 9 PCs containts an 81.2% of the total information. The first factorial map conserves a 38.1% of the total information. For this reason, we will also extend the analysis to the third principal component. The projection subspace formed by the first 3 PCs conserves a 47.4% of original information. Even if this is far from the 80%, with this three axis the phenomenon can be properly described at the level of interpretation. Whenever this PCs should be used in further predictive models, the remaining PC4-PC9 should be also included and properly interpreted for modelling explainability purposes.

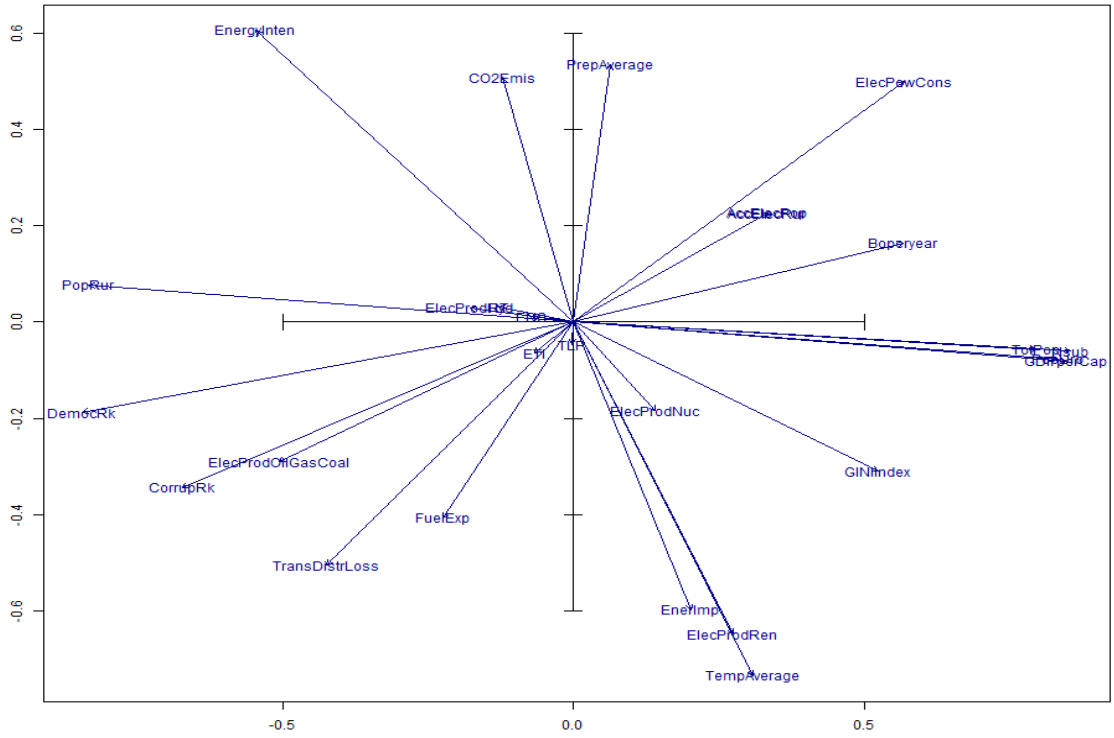


Figure 26. First and Second PCs represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PC

We will represent two-dimensional maps, in which the axes are the PCs. These axes show how variables interact among them and how they are connected, profiling the behaviour of the failures and blackouts of the system, and finally, of the countries where they take place.

We will explore different factorial maps by using different PCs in the factorial map axes.

Figure 26 shows the two first principal components and the projection of all active numerical variables.The arrows that are closer to the axis represent the variables with more weight in the component, i.e., with major contribution to the formation of the axis. We can distinguish a strong association between demographical variables as the total amount of population or the percentage of urban population and the economic indicators of the country (*GDPperCap)* and the first factorial component. The first axis divides the countries into more and less developed ones, placing the more developed countries (with higher total population, higher concentration in urban nucleus , higher GDP per capita and smaller level of rural population) into the right part of the factorial map. The vertical axis is more influenced by the climatic conditions and the sources of energy used. It shows that higher temperatures (*TempAverage*) are associated with less precipitation (*PrepAverage*) and, higher energy production from renewable sources (*ElecProdRen*). Also the energy imports increase with temperature and lack of water (*EnerImp*). As clean production is more important and part of electricity is produced out of country borders, the CO_2 emissions are smaller. Some weak association with lower intensity (*EnerInt*) and power consumption (*ElecPowCons*) is also shown. That is to say, that 2nd PC shows a relationshipbetween the climatic conditions of the country and the energy matrices of production which concentrate in clean energies for warmer countries. Consequently, this is associated with more efficient systems and higher air quality.

Next step on the process consists on projecting the categorical variables in the *MEDB* to the factorial plane. These supplementary variables will contribute to enrich the interpretation of the results of the PCA. The categorical variables included are shown in Table 4:

Year	Climate	ElecTrans	Nuclear	EU
Month	Island	Eleccome	OCDE	
Country	ElecDistr	RegPrice	RatParis	
Reason	ElecGen	Intercon	Government	

Table 4.Qualitative variables added to the first PCA for the MEDB used as illustrative and complementary data

Figure 27 shows all of these variables but Reason and *RatParis*. These variables project all modalities near the center of coordinates of the map, indicating no contribution to the represented factorial components. In Appendix D the map containing all the qualitative variables projected

can be seen. In Figure 27 only relevant qualitative variables for the first factorial plane are projected.

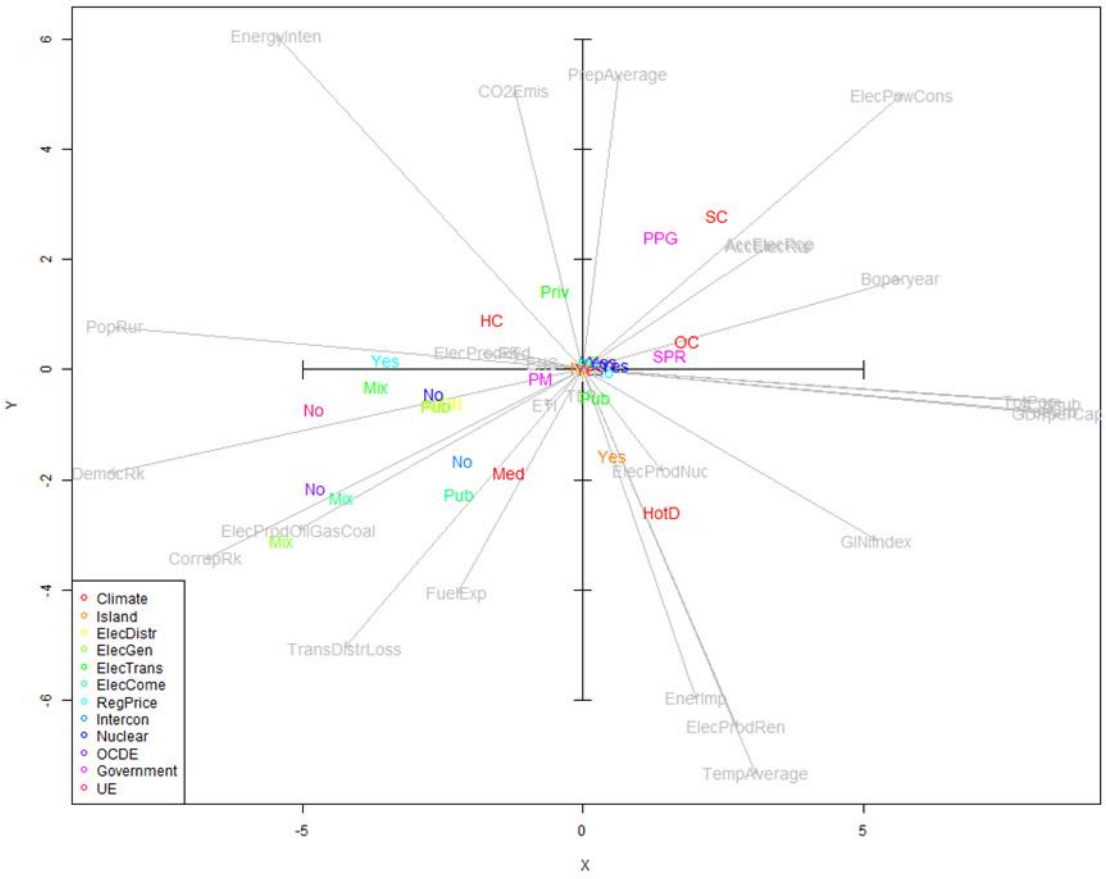


Figure 27. Quantitative active and Qualitative illustrative variables considered for the PCA. First and Second PCs are complemented by the qualitative variables of the MEDB showed in Table 4

Climate				
Humid continental (HC)	Hot Desert (HD)	Subartic continental (SC)	Oceanic (OC)	Mediterranean (Med)
Island				
Yes		No		
ElectDistr				
Mixed (Mix)	Public (Pub)	Private (Priv)		
ElectGen				
Mixed (Mix)	Public (Pub)	Private (Priv)		
ElectTrans				
Mixed (Mix)	Public (Pub)	Private (Priv)		
ElectCome				
Mixed (Mix)	Public (Pub)	Private (Priv)		
RegPrice				
Yes		No		
Intercon				
Yes		No		
Nuclear				
Yes		No		
OCDE				
Yes		No		
Government				
Parlamentary Monarchie (PM)	Parlamentary Republic (PR)	Presidential with Parlamentary Government (PPG)	Semi-presidential Republic (SPR)	
UE				

Table 5. Variables included in the study of the effect of the qualitative variables of the MEDB and their interaction and influence to the subset of active variables that are used to conform the PCs. These are the variables represented in Figure 27.

First interesting thing is that countries placed in the bottom of the factorial map tend to be Islands with Hot Desert climate, this explaining the particular energy policy described before in terms of the numerical variables. These countries, in fact, inherits all properties described for the second factorial component: high temperatures (*TempAverage*), higher energy production from renewable sources (*ElecProdRen*); High energy imports and lack of water (*EnerImp*) and low CO_2 emissions. In fact, this corresponds to the fith pattern identified in the clustering process (Hellenic countries). In addition, countries with Mediterranean climate tend to have a public distribution network and non-regulated prices and tend to have higher fuel exports, and transmission distribution losses while lower power consumptions. This is probably associated

with part of cluster *Central and Mediterranean Europe* countries with mixed public-private energy generation tend also to have mixed commercialization. These countries show higher corruption index, and higher production from fossil sources than other countries and still have lower power consumption. This is associated with patterns from cluster *Balkans* countries. Countries with colder climate have higher power consumption, lower losses, lower production from fossil sources and lower corruption indexes and tend to be associated with higher precipitations and lower temperatures and higher level of development, more concentration of population in urban areas and hold parliamentary presidential governments. This corresponds to cluster *Northern Central Europe* pattern.

Figure 28 shows how countries project on the first factorial map. This image also confirms the alignment between the patterns associated with the clusters previously found and the information extracted from the variable projections in the factorial maps , the horizontal axis places the clusters *Central and Mediterranean Europe* and the *Northern-Central Europe* .

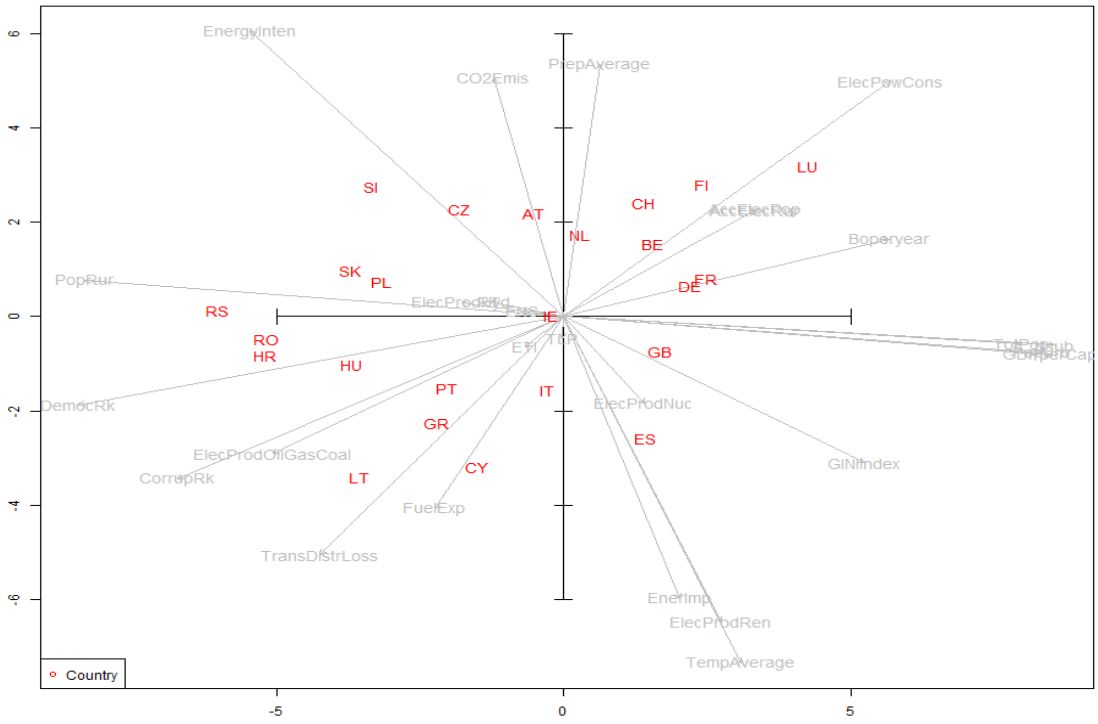


Figure 28. Countries superposed to the representation of the two first PCs, showing the distribution of each of the countries along the axis and its characteristic features.

As we pointed out before, analysing the first two principal components provides only with the 38.1% of the information available in the dataset. This value is far away from the recommendable value of the 80% that will be achieved with the first nine components. To get a wider perspective of the relationships among the variables, we will also analyse the third component. It can be projected with first or second axes. We will analyse the factorial map displaying second and third principal components. Putting aside the first component, which made an economical and demographical division giving few details of the failures and blackouts themselves, we will confront the 3rd and 2nd components as these, may show interesting behaviours and relations among variables.

We repeat the process again obtaining the influence of each variable on the second and third principal components. Figure 29 shows the distribution of the variables along the axis. The second axis is, as it was already seen, directed by variables related to the energy matrix and the characterisation of the electric consume, and the climate of the country where the blackout happened (*ElecPowCons*, *EnergyInten*, *EnerImp* or *TempAverage*). The third component, in the vertical axis, is dominated by the variables *ElecProdHyd*, *Nsub* and *Boperyear*, in the positive sense; and by *ElecProdOilGasCoal* and *ElecProdNuc*, for the negative part. To summarize, Figure 29 suggests that, countries with bigger grids and more events per year also count with

more hydroelectric production. On the contrary, those with smaller grids and fewer failures produce more electricity from conventional sources. An explication to this fact may be the different technologies associated to each of the electricity production means. More conventional and spread sources, as fossil fuels or nuclear fuel, may cause less incidences due to their operation. Nevertheless, this will need a further technical study to refute it and we can not assure it from the data in the map.

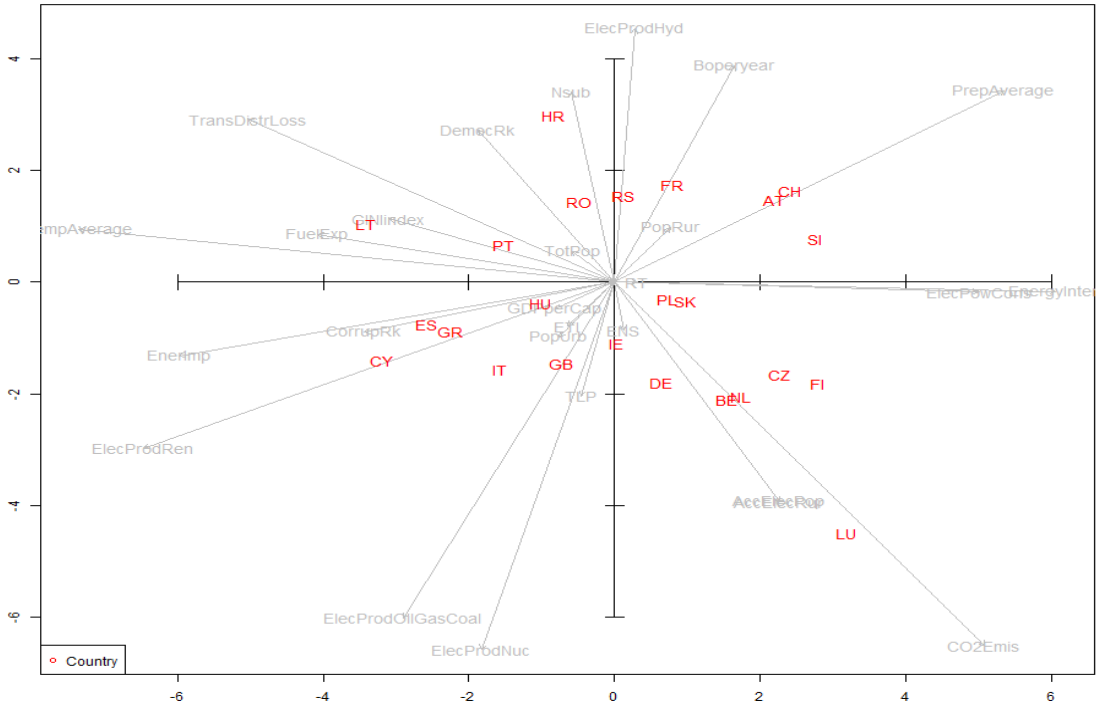


Figure 29. Second and Third PCs represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PCs.

The technical properties of the blackouts have not appeared as influent variables for the principal components. These are headed by the economic characteristics of the countries, as well as by the energy policy of each one. The size of the grid seems to be critical too.

As we are interested in understanding the blackouts phenomenon from a more technical perspective, , we repeat the PCA, this time, only considering as active variables those regarding the technical characteristics of the failures and giving information about the energy policy of the countries. Table 5 shows the variables used in this PCA:

ENS	ElecProdHyd	ElecPowCons	Boperyear
TLP	ElecProdNuc	TransDistrLos	
RT	ElecProdOllGasCoal	FuelExp	
ETI	ElecProdRen	EnergyInten	
CO2Emis	EnerImp	Nsub	

Table 6. Variables used to perform the second PCA for the MEDB. For this second analysis, only the technical variables are included actively in the study.

This time, in order to achieve 80% of the global total variance, we need to keep 8 PCs (83.5% of the total information).

The results of representing the first factorial plane are shown in Figure 30. We find that the horizontal axis, corresponding to the first PC, is dominated by the number of blackouts per year in the positive sense, and by the electric production from fossil fuels in the negative. In the positive part, *Nsub* and *ElecPowCons* are important factors too. The vertical axis is dominated by the energy imports and the electric production from renewable sources in the negative sense, and by the energy intensity and the CO_2 emissions in the positive. Again, the conclusion seems to be that countries with bigger grids have more incidences per year, as they also consume more electric power. France, Austria and Germany seem to behave in this way. Oppositely, countries with more electric production from fossil fuels, as Macedonia or Romania, have smaller grids and less number of failures in a year. Electric production from conventional sources, as fossil fuels, is again associated with less problems in the grids, turning to be this source the most reliable one. Countries whose energy intensities are higher also emit more gases, import less energy from abroad and produce less clean electricity. Bigger or less efficient consumes together with lower purchases of energy from other countries and less clean production, are associated with dirtier atmospheres and more harmful emissions. Czech Republic is a good example in this sense.

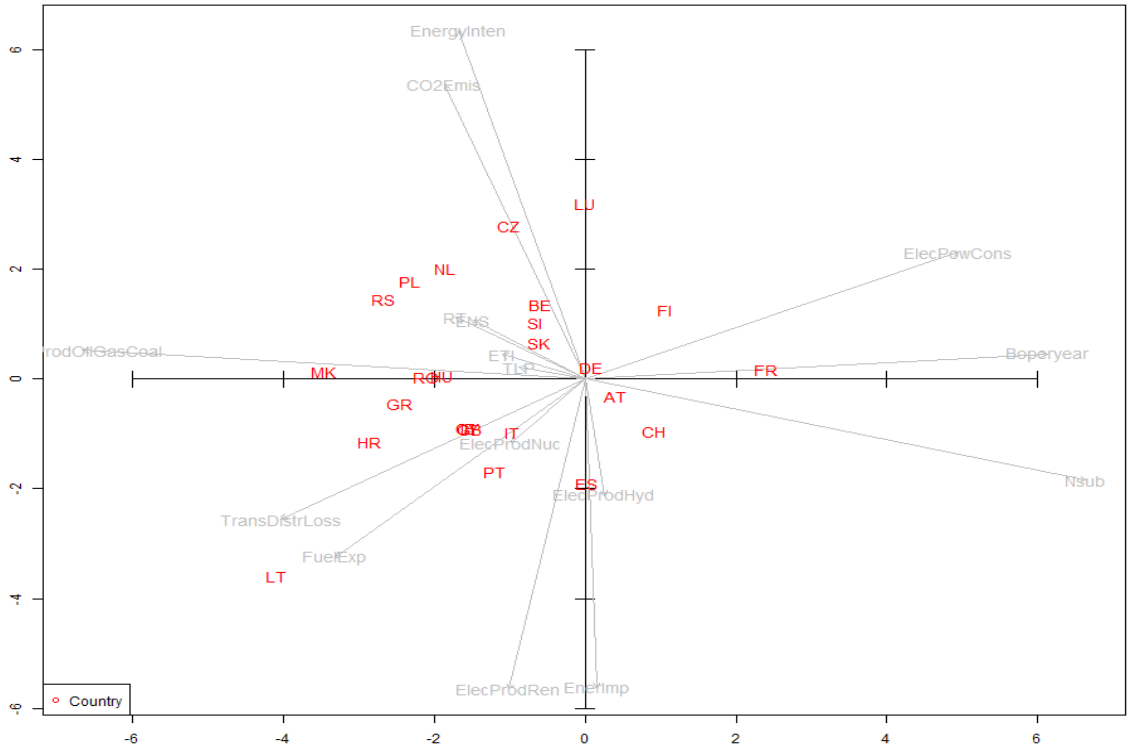


Figure 30. First and Second PCs are represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PCs. Countries are superposed to the representation of the two first PCs, showing the distribution of each of the countries along the axis and its characteristic features.

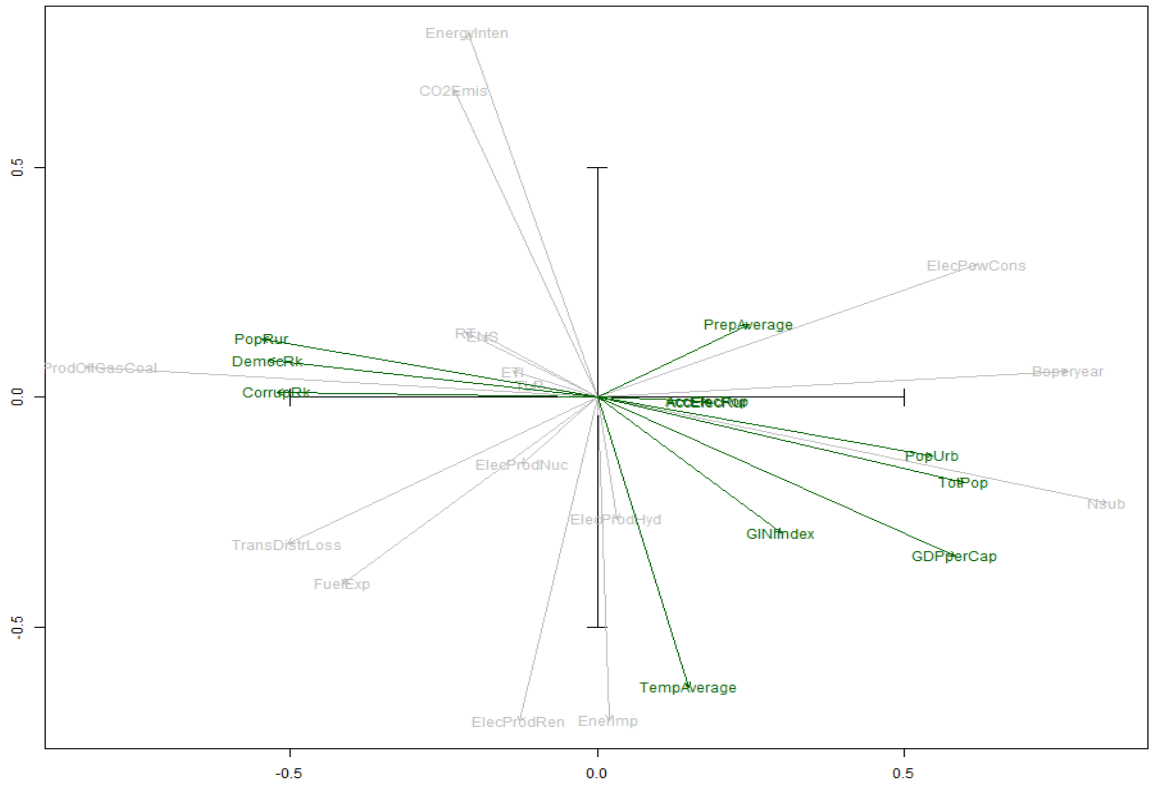


Figure 31. Variables used as active ones for the PCA are represented in grey. Dark green arrows represent the variables of the MEDB that have not been included in the active group of variables, used to build the PCA. The relationship among the illustrative variables and the axis is shown in terms of orientation and proximity to them.

In Figure 31 we add to the representation the economic and climatic variables removed before as illustrative variables. The distribution agree with the conclusions stated. Countries with bigger grids and more failures (Germany, France, etc) have bigger urban and total populations as well as higher GDP per capita. That is to say, that countries with more economic and demographic development in Europe, embrace the bigger transmission networks. Countries using more fossil fuel as electric source, as Macedonia or Romania, tend to be more corrupted and less democratic and have bigger levels of rural population. Therefore, less developed countries in terms of policy and economy are also less developed in terms of energy innovation. Related to the vertical axis, countries with higher average temperatures are the ones producing more electricity from renewable sources.

Next, we confront the first and fourth PC (Figure 32). Even if we previously substituted the first PC by the third one in the representation arguing that, as it talked about economic and demographic factors it was not relevant for the study of the electric failures, once we have removed the economic and demographic variables from the "actives" subset, is interesting to

study it again. This way we can better appreciate the behaviour and direction of the technical variables that are now heading the 1st PC. Nevertheless, this time we can appreciate the influence of some of the technical parameters of the blackouts in the vertical axis. This axis infers that, blackouts lasting more (*RT*) are the ones which does not supply bigger amounts of energy (*ENS*) and take more time to be fixed and to have the service restored (*ETI*). Looking at this axis, we realize that Cyprus is the closest country to it. This is reasonable if we remember that Cyprus registered the biggest and longest blackouts in the *MEDB*.

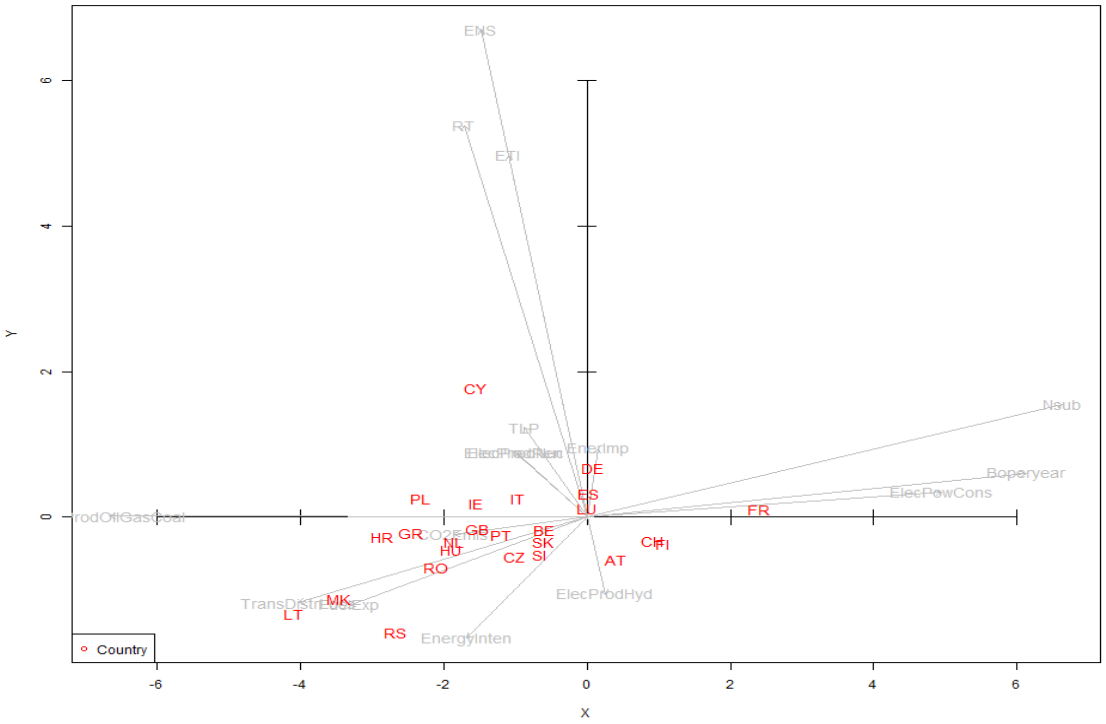


Figure 32.First and Fourth PCs are represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PCs. Countries are superposed to the representation of the two first PCs, showing their distribution along the axis and its characteristic features.

After developing the ACP for the *MEDB* and using several ways to perform it, we conclude that blackouts are more frequent in bigger grids and in countries that consume more power electricity. On the contrary, countries with smaller grids and more important conventional energy sources, as fossil fuels, present less problems in form of blackouts in absolute terms. If these blackouts are more intense, in the way that they last more, they produce a bigger lack of energy and take longer times to return to the normal activity. Patterns showed by the previous maps (Figures 30 and 32) confirm the behaviour of the different clusters generated from the *Country* dataset.

Nevertheless, the *MEDB* is created by adding information of the country in which a failure happened. This means that countries with bigger grids, and consequently, with more incidences appear more in the dataset. This fact bias the results of the study, because economic, policy or climatic variables of these countries are more frequent and have more weights in the PCA. With the aim of better understanding the relationship between the failures in a country and the characterization of the country itself, we proceed to repeat the PCA but this time using the *CDB* as the data matrix to be analysed.

4.3.2 Principal Components Analysis for the Country dataset

Country dataset contains several variables referring to the technical characteristics of the blackouts, as the *MaxETI* or *ENSAverage*, as well as variables related to the country grid and it performance (*Nsub*, *Bonorm*, etc.). Table 6 contains the variables we use this time to build the subset of "active" variables:

CO2Emis	ElecProdOilGasCoal	GDPperCap	PopUrb	PrepAverage	NsubPop
AccElecPop	ElecProdRen	GINIindex	EnergyInten	Nsub	BoyearPop
AccElecRur	EnerImp	FuelExp	CorrupRk	Boperyear	MaxETI
ElecProdHyd	ElecPowCons	TotPop	DemocRk	ENSAverage	MinETI
ElecProdNuc	TransDistrLoss	PopRur	TempAverage	ETIAverage	

Table 7. Variables used to perform the first PCA for the CDB. For this first analysis, all the numerical variables in the MEDB are included actively in the study

Repeating the whole process, we obtain that this time, in order to achieve the 80% of the total variance in the dataset with the principal components; we need to count on 10 PCs (Figure 33).

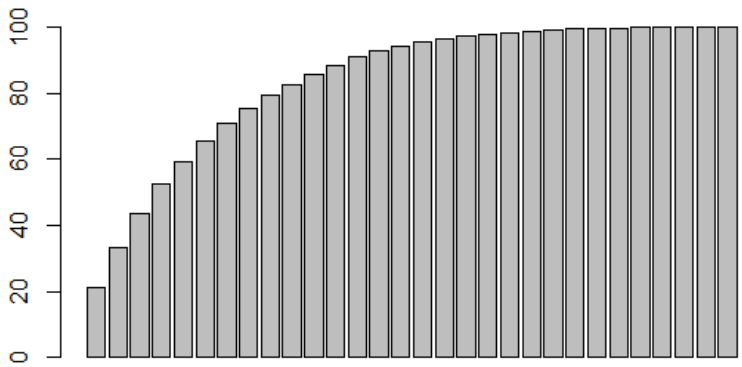


Figure 33. Cumulative percentage of inertia of each of the PCs obtained from the PCA developed using the variables in Table 6

First, we compare the first and second principal components (Figure 34), corresponding to the horizontal and vertical axis respectively. The First principal component confronts countries with more electric power consumption, more urban population and bigger GDP per capita with

countries where rural populations are important, which are consider more corrupted and less democratic and which count on bigger losses in the transmission lines. This structure is also recalling some of the patterns identified in the clustering process. *Eastern Europe* cluster and *Balkans* are well characterized by the negative side of the axis, while *Northern-Central* and *Central and Mediterranean Europe* countries fit better on the right side. The vertical axis exposes that countries with more precipitations produce more energy from hydroelectric sources and suffer from less important electric incidences. On the contrary, countries importing more energy, which have warmer temperatures in average, register longer blackouts, which are also longer to repair. These countries produce more electricity from fossil fuels and have more blackouts per year, taking into account the number of substations that they have. *South Mediterranean* and *Hellenic* countries behave this way, as they show warmer temperatures (specially *Hellenic* countries) and in some cases (as Cyprus or Greece) the longest and most difficult blackouts, in terms of losses. On the other hand, *Northern* and *Central European* countries fit more with the positive part of the axis.

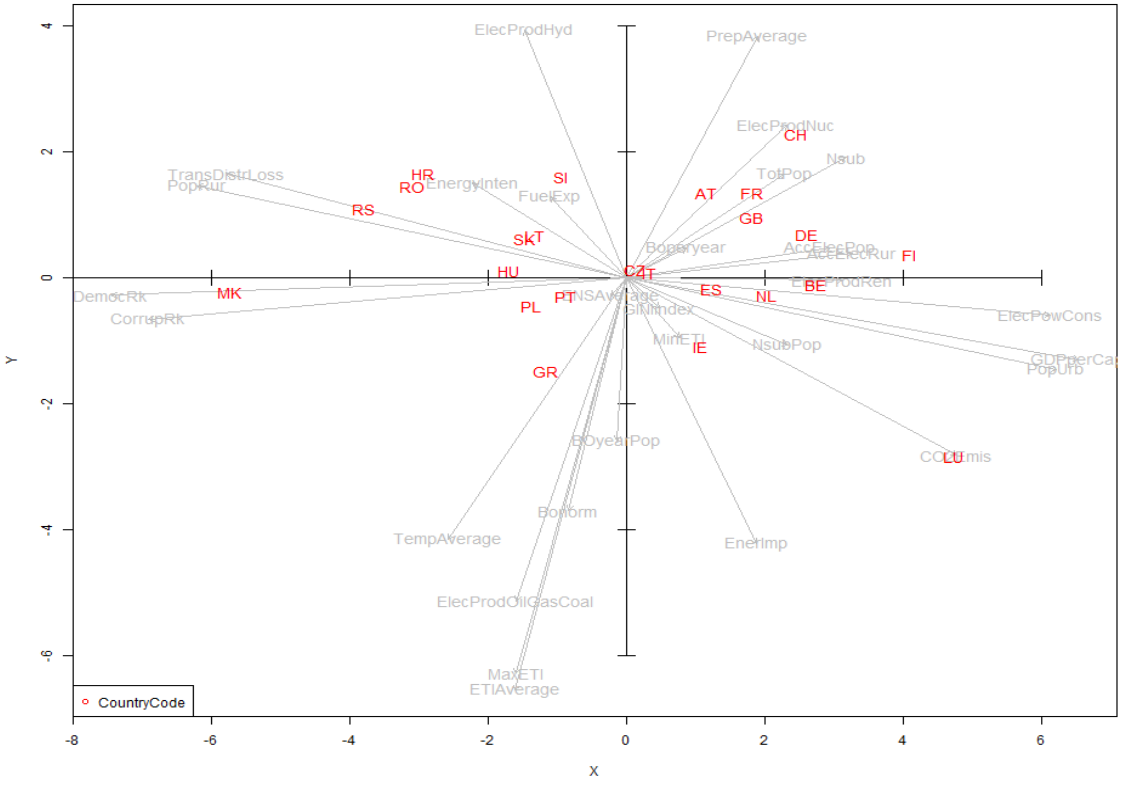


Figure 34. First and Second PCs are represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PCs. Countries are superposed to the representation of the two first PCs, showing the distribution of each of the countries along the axis and its characteristic features.

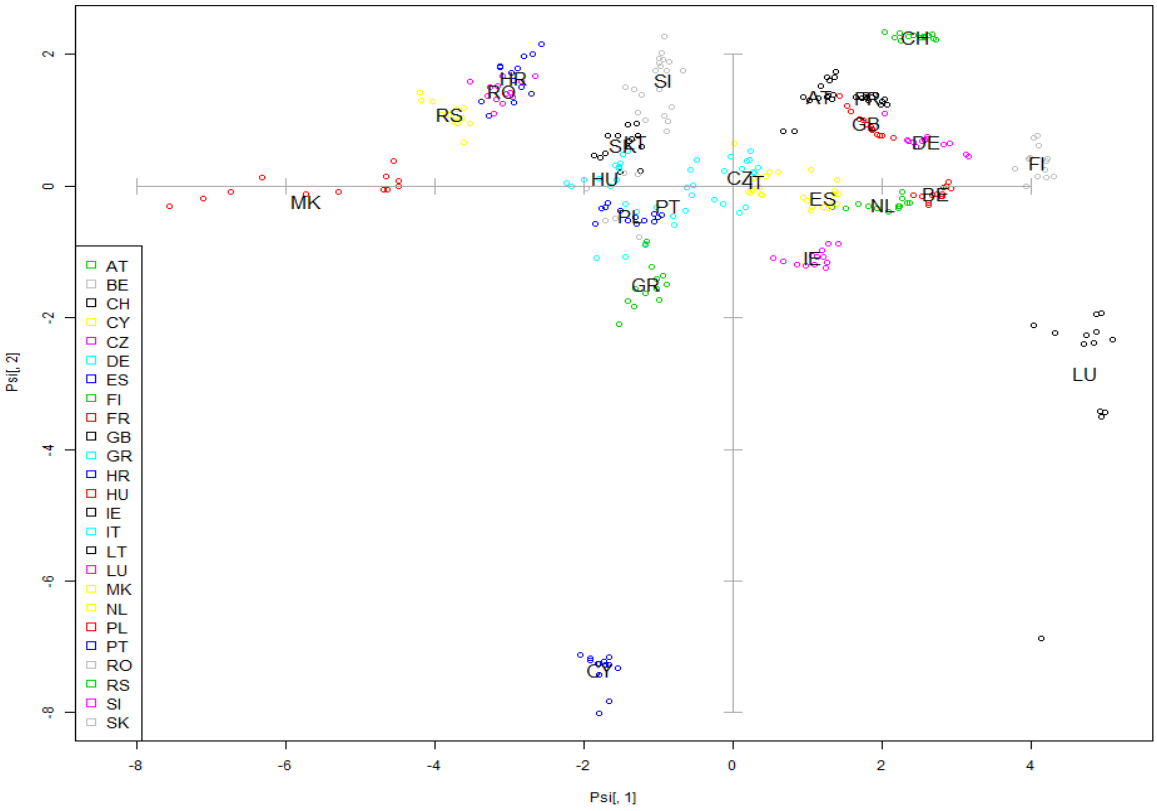


Figure 35. Evolution of the countries of study along the 14 years analyzed regarding the First and Second PC's of the PCA implemented to the CDB.

It is interesting to remark than in this database, each country appears 14 times, one per each year embracing this study. Figure 35 shows the projection of the individuals and the first interesting thing to remark is that all observations corresponding to a same country appear together, so this means that the situation of a country is not showing great changes along the years. Figure 35 shows the evolution of the countries along time. The horizontal axis is an indicator of the level of development of the country, in the sense that bigger positive values mean bigger levels of GDP per capita, CO_2 emissions, electric power consumption and negative values more corruption, less democratic systems and bigger loses during the electric transmission. A movement along the second axis towards the right would imply more precipitations in the country in average terms, less blackouts per year regarding the population, lower medium temperatures and less harmful blackouts in terms of time of interruption of the electric service.

The first visible thing is that the countries do not show big movements along time and keep more or less in similar pattern from the whole study period.

However, one can see that some of the countries evolve horizontally, like Macedonia or Great Britain whereas others evolve vertically like Cyprus or Serbia. This projection do not provide sufficient information to learn whereas time moves upwards or downwards, so we proceed with other two projections that provide additional information to complete this interpretation.

Finally, Figure 36 shows the projection of the variable *Year* in the map. This is useful in order to understand how time evolves in relation with the variables influencing the most the two first principal components. We can see a general transversal evolution of time: first years of the study are place more at the bottom left side of the plane, while, as time goes by, the variable *Year* tends to move to the top-right part of the map. Nevertheless, the trend is not straight, and some kind of zig-zag pattern seems to be followed. In addition, we know that we have measurements from all countries for the 14 years and previous map confirms that observations project grouped by country and not per year. So, more detailed information is required to better understand what happens.

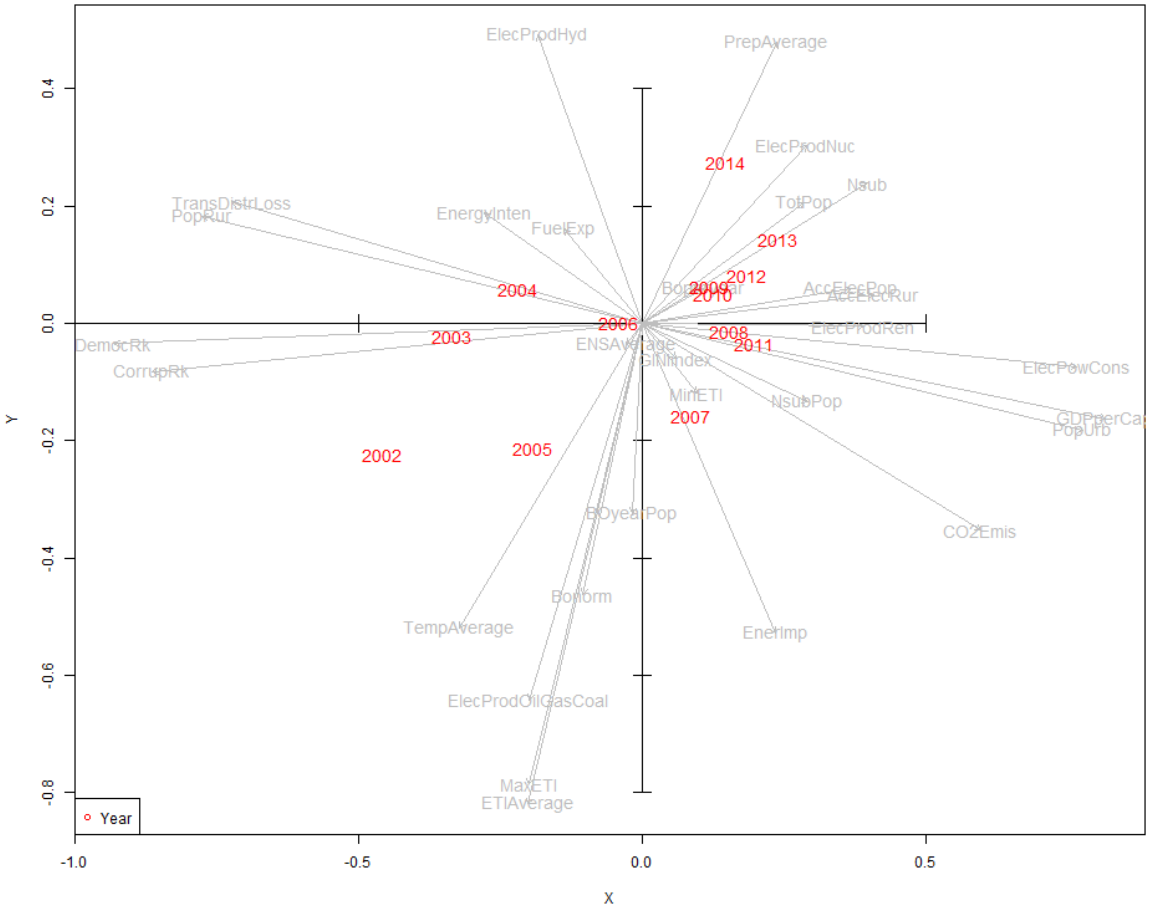


Figure 36. Representation of the evolution of the variable Year along the two Principal Components. This map helps with the interpretation of the variables evolution along time

Figure 35 shows the year of each observation of the projection of individuals, so we can see the date of each point. Combining this figure with the image in Fig 37, we can understand whereas countries evolve along time towards the right or the left or the top or the bottom of the map, and complete the dynamics interpretation of the phenomenon, for the several countries. Overall, we can see a trend in the evolution of the countries along time that move towards the right side of the horizontal axis and the top part of the vertical one. As the map is confusing, partial views are shown in Figures 38 to 41. They confirm that almost all the countries behave this way, being the most important exceptions Cyprus, which evolves in the negative sense of the vertical axis, and Luxembourg and Finland that moves along the positive part of it. Cyprus is increasing the quantity of electricity produced from fossil fuels and registering longer blackouts as time goes by. Countries as Slovak Republic, Lithuania and Slovenia show a similar pattern, but less pronounced in terms of verticality. Luxembourg and Finland register more precipitations and blackouts per year in terms of mean population. There are countries, as Macedonia, Hungary, Belgium, Netherlands, Ireland and Spain that move mainly along the horizontal, positive axis. This means a tendency to more renewable production of electricity and power consumption, and better levels of democracy and corruption as years pass. Germany, France and Great Britain are the sole countries that move positively along the horizontal axis but negatively in the vertical one. This implies bigger CO_2 emissions, energy imports and urban populations. The rest of the countries (i.e. Romania, Switzerland, Italy, Austria, Portugal, Greece…) describe a transversal movement through the positive sense of both axis, with bigger average precipitations values, more production from nuclear sources of energy and total populations.

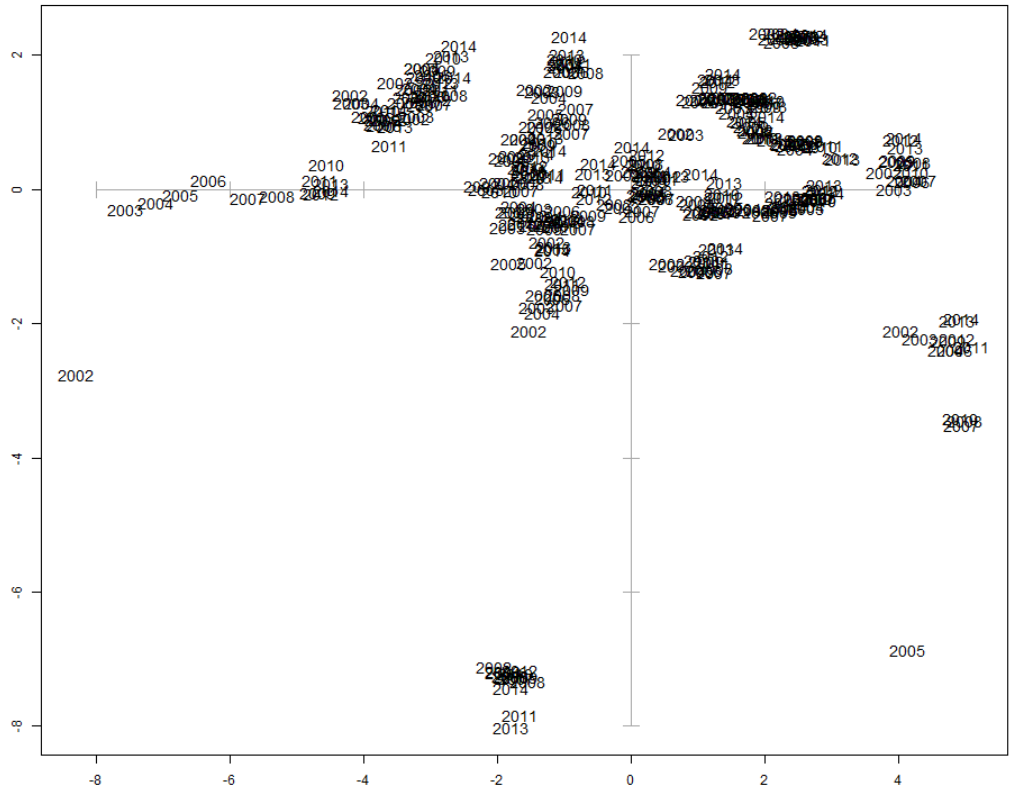


Figure 37. Year evolution for each of the countries in CDB. Every year is represented by one point in the map for every country.

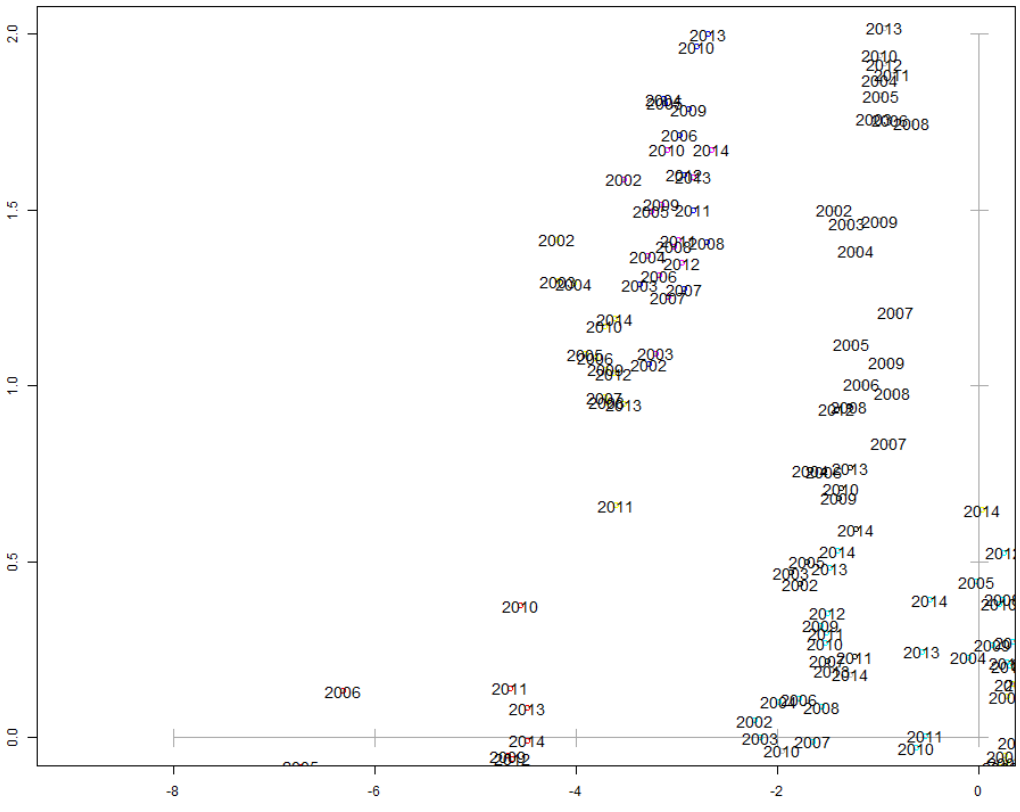


Figure 38. First quadrant of the Year evolution for each of the countries in CDB

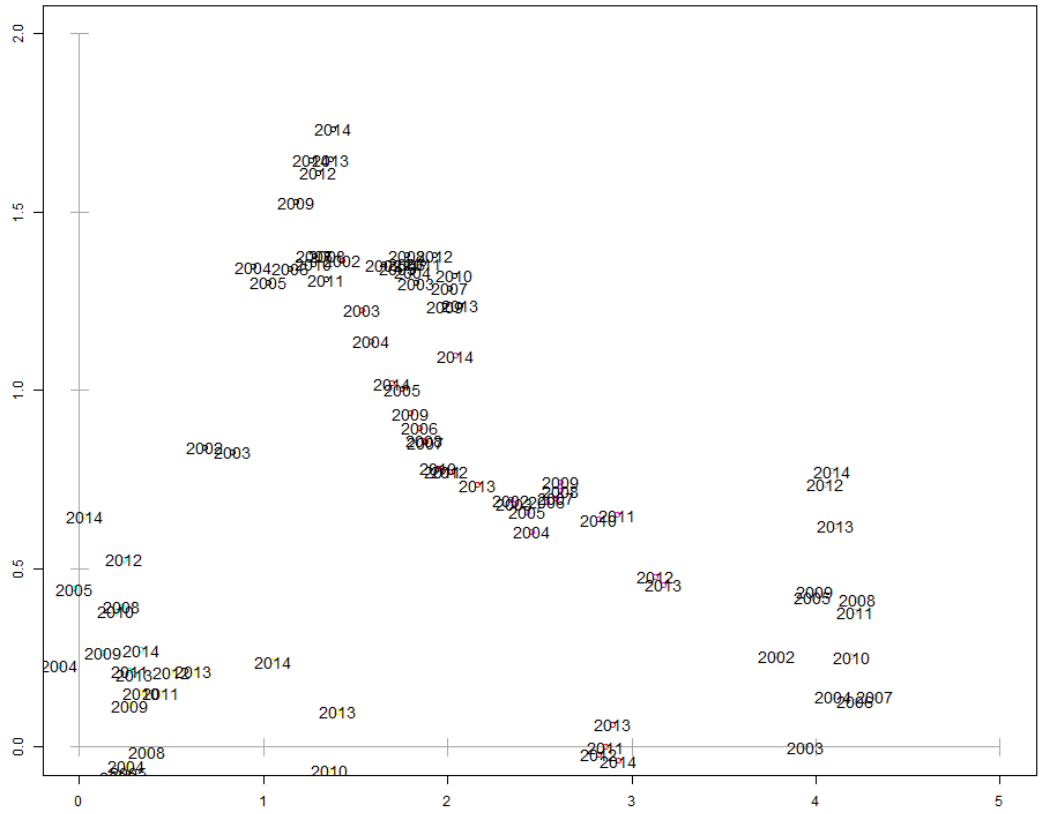


Figure 39. Second quadrant of the Year evolution for each of the countries in CDB

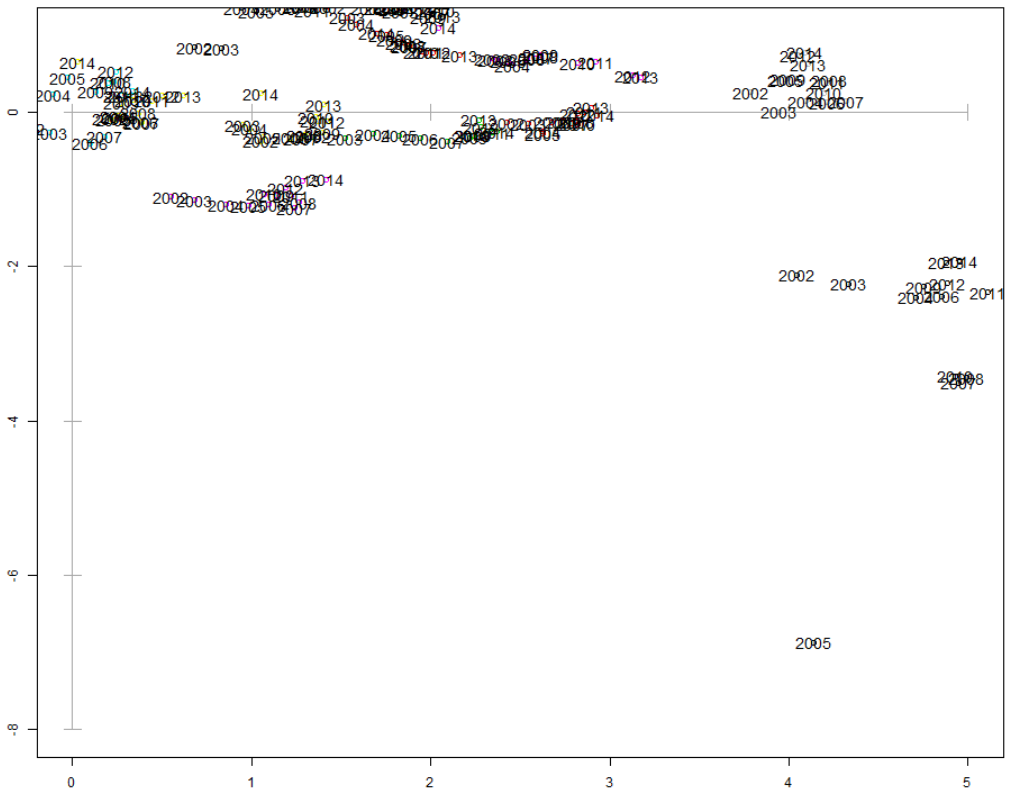


Figure 40. Third quadrant of the Year evolution for each of the countries in CDB

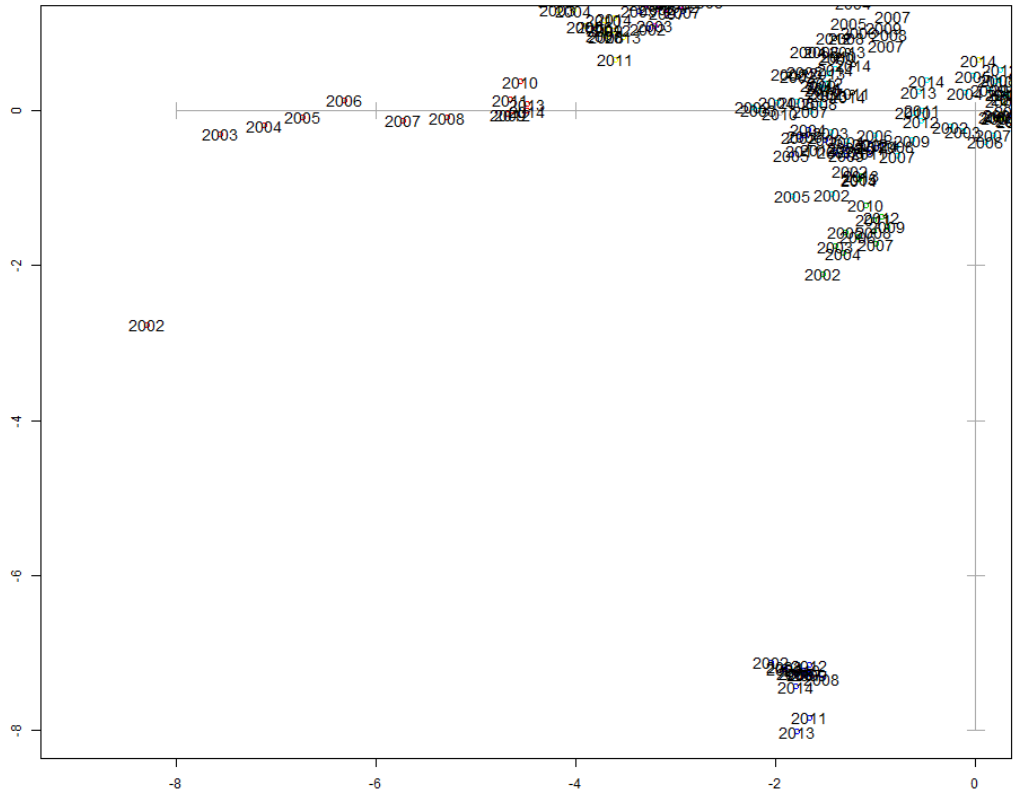


Figure 41. Fourth quadrant of the Year evolution for each of the countries in CDB

The third PC, represented in the vertical axis in the Figure 42, shows again the relationship between hydroelectric production and average precipitation. What is interesting here is the presence of the variable *EnergyInten* in the same direction. That means that even counting with less pollutant sources of electricity, *Eastern* and *Balkans* countries are less efficient in their consumption. On the other side, countries with higher GINI indexes and bigger populations present more failures per year in absolute terms. That is consistent with the conclusion obtained before; more developed countries have bigger electric grids and consequently, more blackouts in absolute terms.

As it happened in the PCA developed with the *MEDB*, the first axis divide the countries in terms of economic development and demography. The second axis considers both climatic and energy policy characteristics. This behaviour is the same we found when analysing the 1st and 2nd PC for the *MEDB*. The difference lies on the fact that the equivalent time of interruption of the electric service due to a blackout is influencing the second component. The 3rd PC indicates that more developed countries, in terms of equity and population, register more incidences as they count with bigger grids.

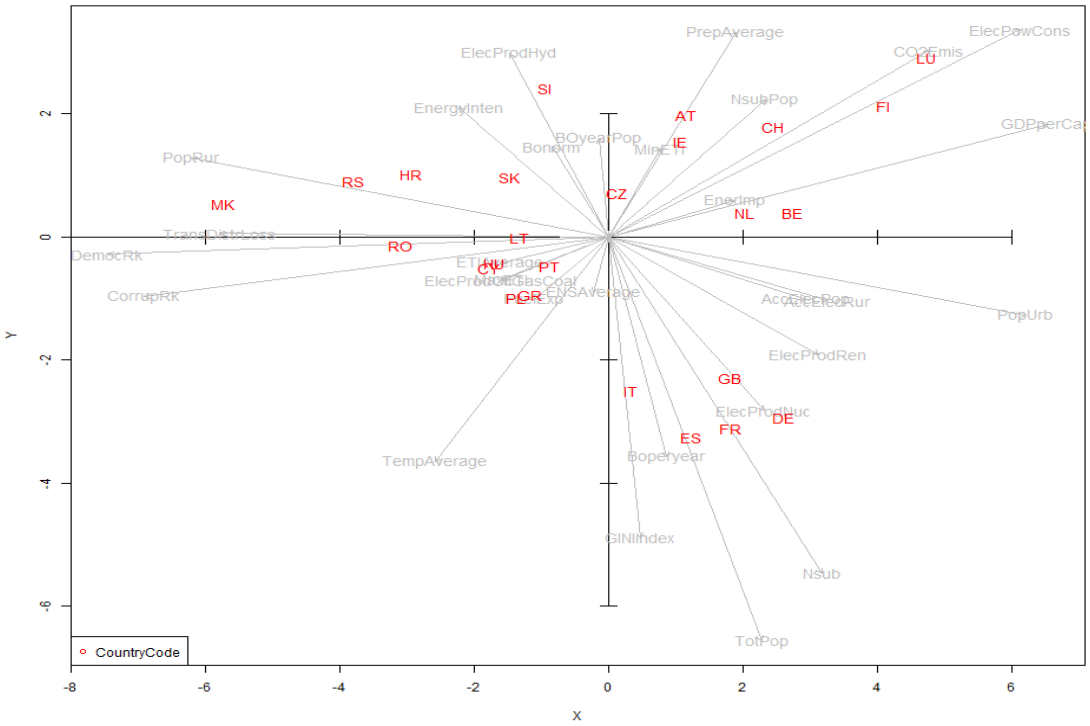


Figure 42. First and Third PCs are represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PCs. Countries are superposed to the representation of the two first PCs, showing the distribution of each of the countries along the axis and its characteristic features

The Principal Components Analysis performed for both *MEDB* and *CDB*, brings some general conclusions. Countries differ between them in terms of population, economy and energy policy and these variables are the ones that actually characterize them. The presence of failures in the electric system is related, basically, with the size of the grid. Oher types of variables do not have a direct relation with their appearance. Countries that have a more developed electric system, in terms of extension and number of substations, are the ones presenting more failures. These countries seem to be also the more developed ones regarding to economy and demography. Finally, there is an association among the duration of the blackout and the losses produced by it, as longer blackouts cause bigger losses of power and energy than shorter ones.

,

4.4 Complex networks analysis

As mentioned in previous sections, power grids are common technological networks, which display substantial non-trivial topological features. Studying national power grids as a CN may help to improve their reliability, connectivity and resilience. Even if the goal of our project is more modest, we do believe that analysing the structure and some of the most descriptive topological parameters of the network will suppose a benefit for our project. Through the study of variables as the degree, the betweenness centrality or the closeness centrality, we characterise the power grid of every country in a different way, just regarding to their network nature. These characteristics provide us with some useful information when profiling the national grids and the electric failures happening in each country.

A network is defined by its number of vertices n, and its number of edges, m. For the European power grid we consider electric substations and transformers as vertices, and transmission lines as edges. The information related to every country grid was obtained from a previous work (M. Rosas-Casals & Solé, 2011). The database provided for this study does not include information relative to three of the analysed countries (Cyprus, Finland and Lithuania) and we will have to deal with the missing data.

Figure 43 shows the visual representation of the topological structures of three country power grids: Greece, Belgium and Spain.

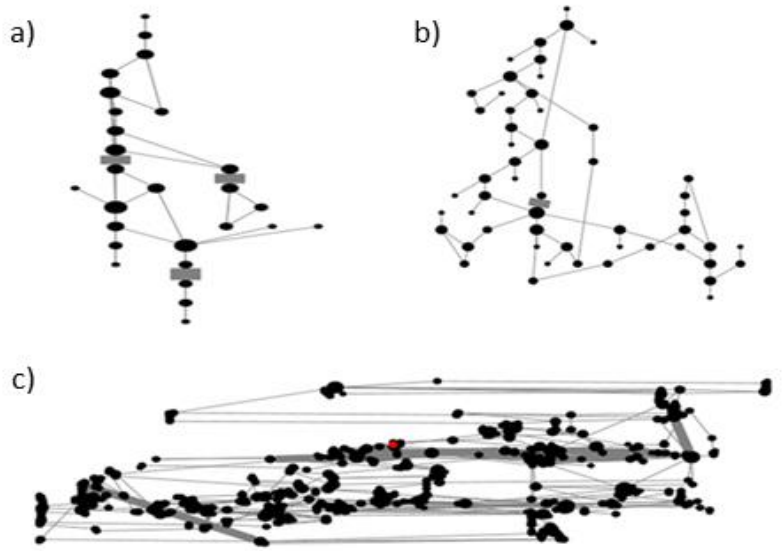


Figure 43. Topological representation Sugiyama of (a) Greece, (b) Belgium and (c) Spain power grids. Substations are represented by dots, which act as nodes, and wires by links, which act as edges of the network. The size of the dot indicates the degree of the node and the width of the link the number of connections that two nodes have in common.

We study some of the topological variables of the network with the aim of characteriseing them. The topological centrality measures (i.e., for each node) that we have used to characterize the structure of networks are the following:

- Number of links connected to a node (i.e., degree of a node).
- Number of shortest paths passing through a node or a link (i.e., betweenness).
- Mean distance to a vertex, averaged over all the vertices in the network (i.e., closeness centrality).

We also include in the study two global measures:

- Shortest path length between two vertices such that no shorter path exists (i.e., average geodesic path).
- Length of the longest geodesic path (i.e., diameter).

In order to make our study more general and useful we transform these topological variables to obtain and work with statistical equivalences, namely: mean, median, maximum and minimum values. We maintain the diameter and the average geodesic path in its original form, being them global measures for the networks. Finally, we have created 12 new variables relative to the network topology of the grids.

Next, we have created a new dataset with the topological information of each power grid. We will refer to it as Country Grids dataset (*GDB*). As we have no detailed information of the evolution of the size of the grid, in terms of addition or removals of substations or wires, during the 14 years analysed, we decide to consider these data constant during the period of study. GDB values are shown in Table 8.

CountryCode	MeanK	MedK	MaxK	MinK	MeanBtw	MedBtw	MaxBtw	MinBtw	MeanClos	MedClos	MaxClos	MinClos	Diameter	AverageGEOPath
CH	2,51	2	11	1	418	148	4168	0	0,014	0,001	1	0,001	19	6,76
SI	2,2	2	5	1	23	11	107	0	0,016	0,0158	0,0238	0,0116	7	3,205
SK	2,419	2	8	1	71	20,73	414	0	0,0056	0,0056	0,0078	0,0039	10	4,2812
RS	2,596	2	9	1	67	14	474	0	0,0057	0,0056	0,0084	0,0035	9	3,8225
RO	2,486	2	7	1	240	105	2328	0	0	0,0018	0,0026	0,0012	11	5,4839
PT	2,552	2	7	1	115	56	681	0	0,0036	0,0036	0,0053	0,002	13	4,9577
PL	2,617	3	9	1	478	216	5509	0	0,0009	0,0009	0,0014	0	16	6,8997
NL	2,111	2	5	1	71	48	267	0	0,0058	0,0057	0,0076	0,0042	11	4,8919
MK	1,846	1	4	1	5	0	20	0	0,1269	0,0476	0,5	0,0345	5	2,0734
LU	1,5	1,5	2	1	1	1	2	0	0,208	0,208	0,25	0,167	3	1,25
CZ	2,432	2	9	1	111	55	795	0	0,0572	0,0034	1	0,0024	9	4,3406
HR	2,132	2	6	1	50	1	279	0	0,037	0,007	0,5	0,005	8	4,0761
BE	2,189	2	5	1	125	100	762	0	0,0034	0,0034	0,0051	0,002	15	5,7038
AT	2,197	2	6	1	181	56	1256	0	0,0024	0,0023	0,0034	0,0016	14	6,0928
HU	2,279	2	9	1	63	7	400	0	0,0522	0,006	1	0,0042	9	4,2124
FR	2,667	2	14	1	2905	664	76502	0	0,0011	0,0002	0,1667	0,0001	26	9,6553
DE	2,499	2	9	1	2400	710	28593	0	0,0136	0,0002	1	0,0001	31	11,9773
IT	2,676	2	8	1	1060	339	11757	0	0,0162	0,0004	1	0,0002	28	9,9291
ES	2,848	2	14	1	1726	403	40499	0	0,0003	0,0003	0,0004	0,0002	18	8,2822
GB	2,526	2,5	5	1	1495	671	11335	0	0,0003	0,0003	0,0004	0,0002	32	12,2353
CY														
FI														
GR	2,444	2	5	1	42	25	131	0	0,0096	0,0095	0,0133	0	10	4,0658
IE	2,516	2	7	1	31	27	177	0	0,0748	0,011	1	0,0068	8	3,2379
LT														

Table 8. Variables and values composing the GDB. Blank spaces represent missing data.

4.5 Principal Components Analysis including complex network characteristics

Our last step is to develop a new PCA, this time considering the variables obtained from the complex analysis study. As it is shown in Table 9, 3 out of the 25 countries studied do not have values for the network variables analysed. The reason is that the database used to build the networks does not contain information for Cyprus, Finland and Lithuania.

With the aim of obtaining a complete dataset, we implement again the MIMMI method. First step is to join *GDB* and *CDB* in order to count on a bigger dataset containing useful information about the countries. The variables included in *CDB* are used to create clusters of countries and calculate the group mean for each of the variables in the *GDB*. As happened in the first missing data imputation, the appropriate number of clusters to choose is 6. That is logical as we are creating the groups from more or less the same variables than before. Figure 44 shows the cluster distribution. We substitute the missing values by the group mean values of the cluster where the country belongs to. Once we count with all the values of the *GDB* we join this dataset to the *CDB*. We will refer to this new dataset as *CDB+*.

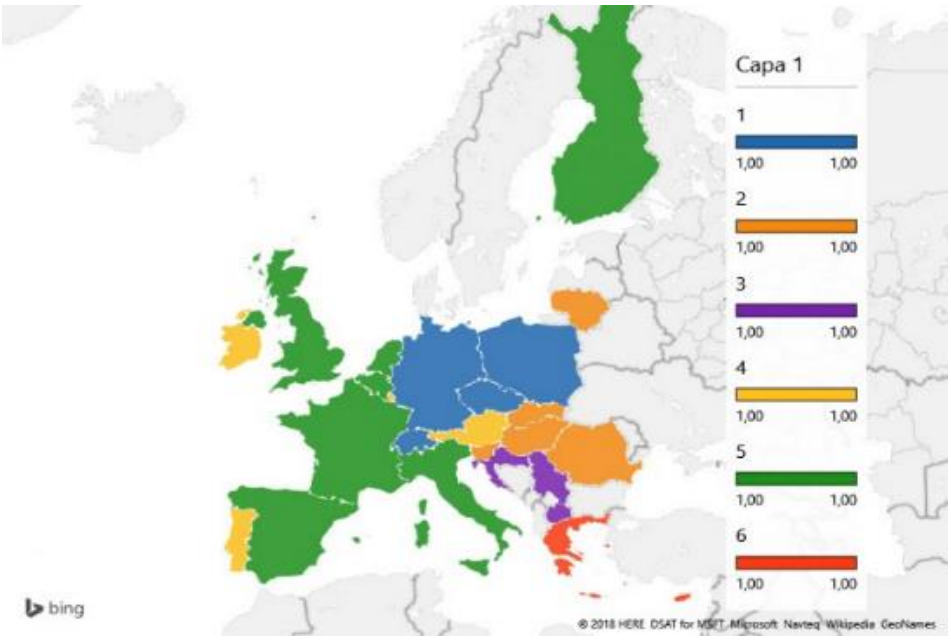


Figure 44. Clusters created after the implementation of the MIMMI method used to impute the missing data of the GDB. A mean value is calculated for each of the 6 groups and the 14 years studied. Missing values are substituted by the mean value of their group and year.

We repeat now the PCA analysis using *CDB+* as our source of information. The main objective of this new PCA is to compare the results with the ones arisen from the previous PCA implemented (i.e., the one related to the *CDB*). In this study, we used all the variables in the dataset for the creation of the PCs. Now we work only with those variables related to the technical

characteristics of the blackouts and the energy policy of the countries, adding the variables obtained from the complex networks analysis, using them as active variables to build the PCs. The aim is to simplify the study and produce some material in which economic and demographic variables do not mask the results using them as active variables to build the PCs. The variables included in the study are shown in Table 10.

CO2Emis	EnerImp	Nsub	BoyearPop	MaxK	MinBtw	Diameter
ElecProdHyd	ElecPowCons	Boperyear	MaxETI	MinK	MeanClos	AverageGEOPath
ElecProdNuc	TransDistrLoss	ENSAverage	MinETI	MeanBtw	MedClos	
ElecProdOilGasCoal	FuelExp	ETIAverage	MeanK	MedBtw	MaxClos	
ElecProdRen	EnergyInten	NsubPop	MedK	MaxBtw	MinClos	

Table 9. Variables used to perform the first PCA for the CDB+. For this first analysis, only the technical variables in the MEDB are included actively in the study.

We use these numerical variables as active ones, in order to create the set of principal components. Figure 45 shows the cumulative percentage of inertia added by each additional PC considered. In order to work with the minimum of the 80% of the total variance present at the dataset we need to work with 7 PCs (84.34% of the total information).

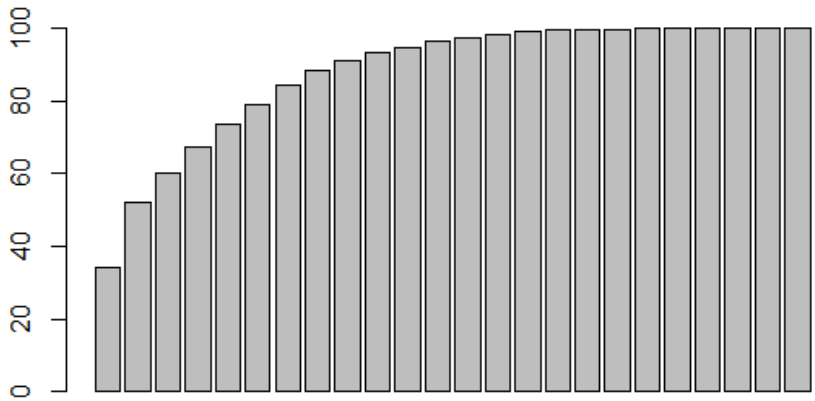


Figure 45. Cumulative percentage of inertia of each of the PCs obtained from the PCA developed using variables in CDB+ (Table 8).

Results for the first and second PCs are shown in Figure 46. We observe how the degree (*K*) is the main influence of the first axis. In the negative sense of it, the maximum degree (*MaxK*) heads the first PC. In addition, the mean and the median degree are quite close to the axis. The average

geodesic path and the diameter have some influence in the axis too. Consequently, bigger degrees of the power grid, in the form of maximum, mean or medium degree, are associated with bigger diameters and short paths in the network. That is to say, grids with more connections (i.e., bigger mean or median degrees) and with presence of much more connected nodes (i.e., maximum degree) are bigger ones, as the longest and the shortest paths to reach one node from another, are longer. Countries with bigger grids are also the ones holding more connections in between nodes or substations. This is how the grids from Great Britain, Spain, Germany or Italy perform. The second PC confronts countries with bigger losses during the distribution of electricity (Croatia, Serbia, Lithuania…) with countries that produce and consume more energy from renewable resources. This can be caused by the fact that, many times, renewable energies are used nearer the production points, while other types of energy are transported through long distances, suffering from higher losses along the way. We realize that countries with smaller and more centralized grids, as a result of their fewer connections, and with more losses of electricity during the distribution, are all quite close distributed along the axis, showing they are more similar. In opposition, countries with bigger and more connected grids differ more in their behaviour and characteristics, spreading more in the graphic.

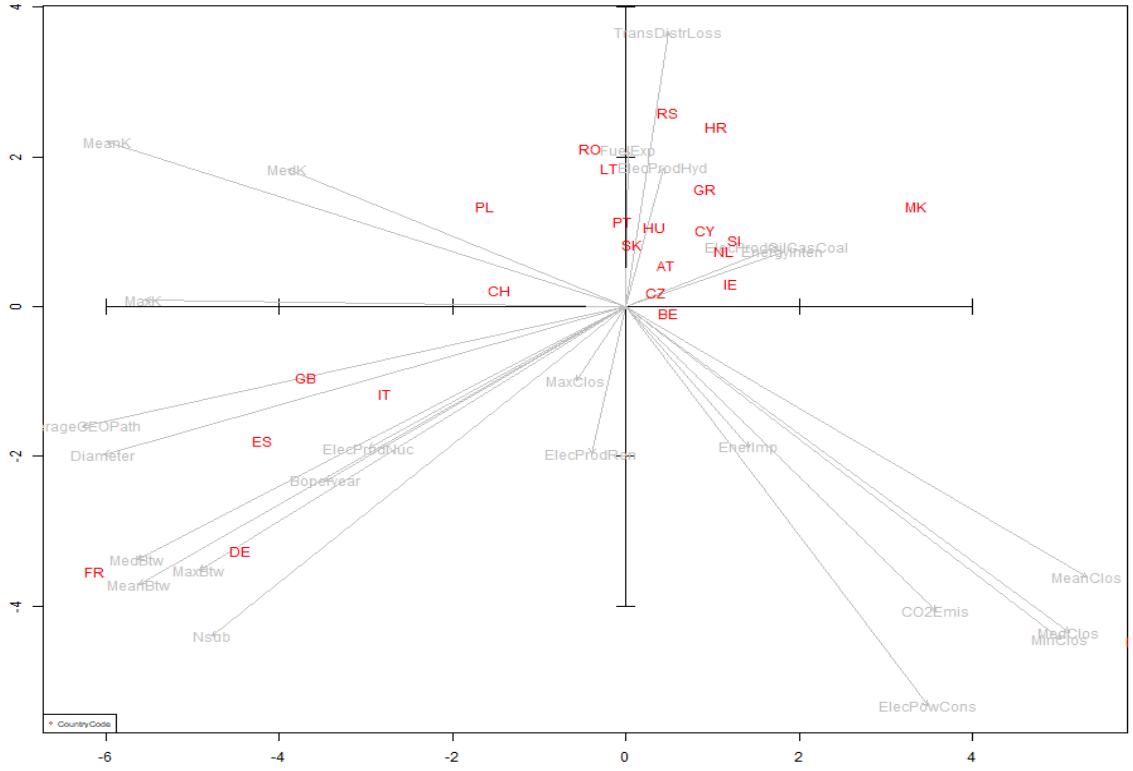


Figure 46. First and Second PCs are represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PCs. Countries are superposed to the representation of the two first PCs, showing the distribution of each of the countries along the axis and its characteristic features.

When we compare second and third axes (Figure 47), we realize that countries with more power consumption, bigger grids (*NSub*) and more blackouts per year in absolute terms, are the ones showing bigger influence of the betweenness and the closeness centralities of the grid, as all of these variables are placed in left side of the horizontal axis (2nd PC). This is also a consequence of the size of the grid. Having higher values of betweenness and closeness centralities mean that, in order to arrive to a given node it is necessary to go through more nodes. That is to say that the diameter (an indirect measure of the size of the grid) is higher. Spain, France or Germany power grids are again the ones following this trend. In opposition, countries as Croatia, Serbia or Romania, with bigger losses during the electricity distribution and more imports of fuel, are the ones with smaller networks. This means that, grids with more connexions diminish the distance between nodes, and as the losses are directly proportional to this distance, they become more efficient.

The third component, represented here by the vertical axis, seems to divide countries with more production from fossil fuels and countries with higher maximum degrees of the network and higher energy intensities. As we find out before, more production from fossil fuels are related to more electric failures, when talking about normalized measures by population and number of substations. And this is opposite to the size of the grid (countries with bigger grids show a more flexible and diversified electricity matrix), which is correlated to more blackouts in absolute terms.

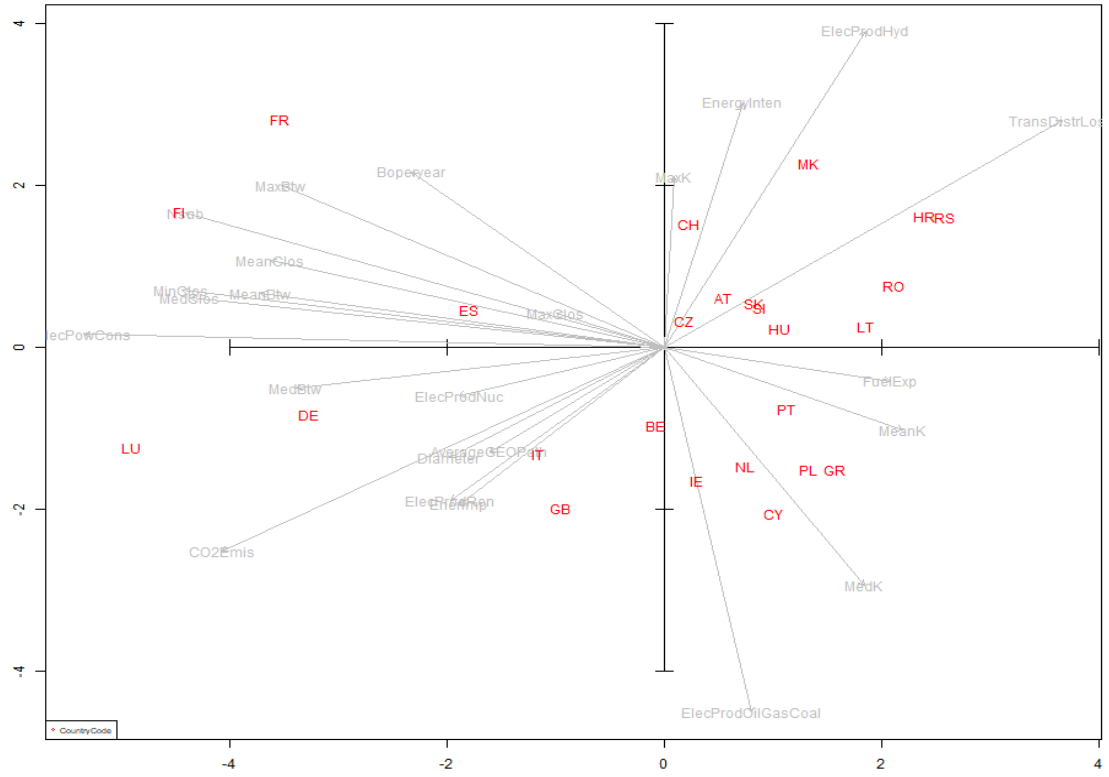


Figure 47. Second and Third PCs are represented in the horizontal and vertical axis respectively. Arrows represent the variables influencing and conditioning each of the PCs. Countries are superposed to the representation of the two first PCs, showing the distribution of each of the countries along the axis and its characteristic features.

The conclusion obtained from this last analysis is that the topological parameters of the power grids are some of the most decisive variables in the dataset. The 1st and 2nd axis note that grids with bigger degrees of the nodes and bigger betweenness and closeness centralities are the ones consuming more electric power and with bigger grids, both in topological terms and number of substations. The fact of counting with bigger degrees and centralities is translated into more connected grids and less losses during transmission. Substations, understood here as nodes, have more connexions with other substations and they lay on the path of more wires to reach their destiny, which implies, decentralized networks. Countries as Germany, France or Spain seem to be good examples of countries in which the power grid is large and not depending on some central nodes. Lithuania, Portugal, Romania or Hungary seem to be good examples of smaller grids but more centralized and dependable on some few substations.

5 Discussion

Reviewing the work done, we are able to suggest some improvements and changes that could have been interesting to develop if possible. The main issue has been the lack of some of the data we needed to develop the study. We did not have the information relating some of the studied countries in the first database used (ENTSO-E database), as well as we did not count with some of the values of the variables of studied for some countries and years. This fact made us implement the MIMMI method and other methods to deal with the missing data. As a result, the study starts from biased data and the quality and reliability of the conclusions could improve if we had all the data available.

First, datasets were created using open data sources and a previous database from ENTSO-E. The problem with these sources of information is that countries are free to provide with the information they want. Sometimes countries hide or even don't give any information, when this can cause a problem or they do not feel comfortable with it. In addition, in order to obtain missing values from the data, which came from the lack of information about some years or wholes countries, may have biased the data. Implementing the MIMMI method to deal with missing values assumes that similar countries behave equally, and this is not always true. The perfect situation would have been to work with fully reliable data nevertheless it was not completely possible.

Regarding data used for the study, it would have been useful to count with more data related to technical issues about blackouts, energy policy and/or failures in the system due to other causes. Gathering data from open sources make this objective difficult, as many of this kind of information are private and belong to utilities, companies and/or governmental entities.

On the other hand, given the duration of the project and difficulties in gathering data, previously commented, the omplex Network analysis implemented must have been a limited rather superficial one. Nevertheless, it would be useful and interesting to find out more data about the features of the network and further studies can follow this direction. CN studies have been implemented during the last years to study the resilience of the Power grids, their tolerance to the failures or attacks, etc. Analysing the kind of network we are dealing with (i.e. small-world grids, random grids, etc.) will help to understand their performance and to predict their conduct.

Although we believe that our results offer a novel approach to understand the relation between these variables, one of the objectives stated at the beginning of the projected, namely to improve our ability to maintain and guarantee the ETPG's resilience, defined by its structural integrity, security of supply and transport efficiency, cannot be achieved. A deeper study into the

characteristics of the Power grid as CNs would help to predict the failures in the systems and to understand how to act in order to improve its reliability.

6 Conclusions

We started this study with the aim of finding some connexions among the economic, climatic and energetic policy of 25 European countries and the occurrence of major failures in their Power grids. Following a pre-stablished methodology we made steps in order to obtain some remarkable conclusions. Through the compilation of information after a deep research phase, we built the two datasets used in the work, namely, *CDB* and *MEDB*. Once all the data was prepared, we characterized the studied variables in statistical terms. As a result of this analysis, we achieve a general vision of the set of variables and their behaviour. We realize that technical variables related to the blackouts show a similar tendency for every country and year: they present small, constant values, besides some exceptions. We observe this behaviour in other variables as the access to electricity of the population, which does not shows any difference in the whole period in any country. To the contrary, variables such as the percentage of electricity generated from each energy source change considerably from year to year and between countries. The conclusion of this statistical basic descriptive analysis is that, even if we are dealing with European countries, which have many features in common and which behave similarly, due to shared policies and history, there are still some dissimilarities in terms of development and in the way countries develop their economy and policy. In addition, the climatic conditions of each country also have a substantial impact in the way they have built up their national identities.

Using this information as a guide, we divided the 25 countries into 5 clusters depending on their defining characteristics. By means of the clustering profiling we realized that variables as the GDP per capita, the total population of the country, the percentage of urban and rural population, the corruption and the democracy level of their institutions, the size of the transmission power grids or the duration of the blackouts registered; are some of the clue variables that make the difference among countries. Most of the countries perform similar when they share the geographical location, the climate or their recent history. For example, *Eastern Europe* group is formed by countries that are located in a close region and which share a recent political history that have marked their manner to develop. *Central and Mediterranean Europe* group have a long tradition of supremacy in Europe and they have always been part of the global economic powers.

The way the power grid has been implemented in the countries, as well as their antiquity and the dimension of the country consumption appear as decisive points in the current characterization of the power grids by groups. Nevertheless, the variables affecting the most to the division in groups are the economic and demographic ones, while technical variables regarding the transmission grids are less decisive in the creation and differentiation of groups, with the exception of the size of the grid, represented by the variable *Nsub*. Following the stablished methodology, we proceeded to implement the PCA using the *MEDB* as source of information. Lately, the same

process was implemented to the *CDB*. Conclusions obtained from both are similar. PCA shows how bigger countries, in terms of population, count with stronger economies and higher equity levels. These are the countries showing more innovation and diversification of their electrical sources, as they use more renewables while at the same time they import energy. They do count on the bigger power networks and the size of these power grids is associated with the number of electric incidences in absolute terms. On the contrary, countries that are less developed in economic and demographic terms follow a tendency to use more conventional sources of energy to produce their electricity. These sources of energy are linked to less amount of blackouts in absolute terms. Nevertheless, regarding the number of blackouts normalized by the total population, we realize that these blackouts are more common and more important in terms of duration and losses. On the other hand, climate seems a decisive criterion related to how each country takes profit of different energy sources.

In general, failures in the transmission systems are not linked to any special feature of the countries, nor economic, climatic or related to their energy policy. The variable affecting the most the number of blackouts in a territory is the size of the grid: the bigger the grid, the higher the number of blackouts. In spite of this fact, if we take into account the population to whom they give service or the number of substations in the systems, we realize that, in proportion, smaller power grids present more failures than bigger ones. In any case, even if failures are not directly influenced by other national characteristics, their technical features are correlated. Longer blackouts cause bigger losses of energy and power in the systems and usually, they take longer times to repair the damages. The amount of power losses during the electricity transmission is more frequent in countries with smaller grids and lower economic development. This fact suggest that, countries with stronger economies have built and improved their electric grid over the years, achieving more reliability and security in the service, which is translated in more efficiency in the transport.

The evolution of the variables along the 14 years of study is stable and we cannot observe huge changes in their behaviour. In general, all the countries tend to better democratic systems, bigger total populations and higher GDP levels per capita. Countries also tend to register less important blackouts in terms of duration. Nevertheless, as we explained in Section 4.3.2 there are some countries that do not follow the same pattern and perform differently.

Our last phase of the study was devoted to a Complex Network analysis of the European Power grids. We created a new dataset, *CDB+*, which included information about the topological parameters of the network of each country. Classic defining parameters as the mean degree or the diameter of the grid were studied. Using this dataset we repeated the PCA analysis, this time, only considering the technical and topological parameters of the countries and their power grids. The

conclusions obtained follow the trend previously outlined. Bigger grids in terms of number of substations are the ones with higher degrees, diameter, betweenness and closeness centralities. This fact implies that the network is, in general, more connected and less dependable of some few but important nodes. This fact is related to less power losses in the electricity distribution, what means that, grids with more connexions diminish the distance between nodes, and as the losses are directly proportional to this distance, they become more efficient. That is to say that, countries as Germany, France or Spain, the ones with bigger grids in terms of size, more connectivity in terms of centrality and showing fewer incidences of mean failures by population, have more decentralized and efficient systems.

In summary:

- Defining characteristics of the analysed countries lie on their economic and policy behaviour, more than in their technical structure.
- Even if researching European countries make the differences among them more weak, we are able to find remarkable different features, suggesting that there still exist differences in the way European citizens live their daily lives. Variable influencing more the appearance of major electric events in the countries is the size of their Power grid:. the higher the number of nodes and connections, the more the probability of suffering from failures.
- The evolution of the countries along the period of study is stable and they do not show big changes nor in their characteristics, nor in the typology of their blackouts.
- Nevertheless, there are some variables that are linked to less blackouts or at least, to more extended Power grids, as the diversification of the energy matrix of the country or the decentralization of the system, understood as a complex network. Finally, our analysis suggests that longer failures are usually related to more important losses in the system and longer reparation times.

We hope that a future and deeper study into the characteristics of the Power grid as CNs will help to predict the failures in the systems and to understand how to act in order to improve its reliability.

7 References

Carreras, B. a., Lynch, V. E., Dobson, I., & Newman, D. E. (2002). Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos (Woodbury, N.Y.)*, *12*(4), 985–994. https://doi.org/10.1063/1.1505810

Chateau, F., & Lebart, L. (1996). Assessing Sample Variability in the Visualization Techniques related to Principal Component Analysis : Bootstrap and Alternative Simulation Methods .

Dobson, I., Carreras, B. a, Lynch, V. E., & Newman, D. E. (2007). Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *17*(2), 26103. https://doi.org/doi: 10.1063/1.2737822

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 37–54. https://doi.org/10.1145/240455.240463

Gibert, K. (2014). Mixed intelligent-multivariate missing imputation. *International Journal of Computer Mathematics*, *91*(1), 85–96. https://doi.org/10.1080/00207160.2013.783209

Gibert, K., Sànchez-Marrè, M., & Izquierdo, J. (2016). A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. *AI Communications*, *29*(6), 627–663. https://doi.org/10.3233/AIC-160710

Luo, L., & Rosas-Casals, M. (2015). Correlating empirical data and extended topological measures in power grid networks. *International Journal of Critical Infrastructures*, *11*(1). https://doi.org/10.1504/IJCIS.2015.067396

Newman, M. E. J. (2010). *Networks. An introduction*. Oxford; New York: Oxford University Press.

Pagani, G. A., & Aiello, M. (2013). The Power Grid as a Complex Network: a Survey. *Physica A*.

Rosas-Casals, M., & Solé, R. (2011). Analysis of major failures in Europe's power grid. *International Journal of Electric Power & Energy Systems*.

Rosas-Casals, M., & Solé, R. (2011). Analysis of major failures in Europe's power grid. *International Journal of Electrical Power and Energy Systems*, *33*(3). https://doi.org/10.1016/j.ijepes.2010.11.014

Shlens, J. (2014). A Tutorial on Principal Component Analysis, (February). https://doi.org/10.1.1.115.3503

Solé, R. V., Rosas-Casals, M., Corominas-Murtra, B., & Valverde, S. (2008). Robustness of the European power grids under intentional attacks. *Physical Review E*, *77*(2), 26102. https://doi.org/10.1103/PhysRevE.77.026102

Url, S., & Society, I. B. (1971). A General Coefficient of Similarity and Some of Its Properties, *27*(4), 857–871.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks, *393*(June), 440–442.

8 Appendices

8.1 Appendix A. Basic descriptive analysis for the country dataset

The division of the variables into groups depending on their nature is the one stated in section 4.4.

The basic descriptive analysis will be hold using the following tools:

1. For every numeric variable, a histogram and a boxplot are shown.
2. When possible, regarding to the nature of the variable and its characteristics, a time series plot is shown and also, a plot showing the different values taken by the variable for each year of study.
3. For every categorical variable, a pie of frequencies and a barplot are shown.
4. For every variable of study, statistical indicators are studied as the mean, the median, the typical deviation or the frequency of appearance.

There are 43 variables of study or columns. The variables in the dataset are:

- Country
- Year
- CO2Emis
- AccElecPop
- AccElecRur
- ElecProdHyd
- ElecProdNuc
- ElecProdOilGasCoal
- ElecProdRen
- EnerImp
- ElecPowCons
- TransDistrLoss
- GDPperCap
- GINIindex
- FuelExp
- TotPop
- PopRur
- PopUrb
- EnergyInten
- CorrupRk
- DemocRk
- Climate
- Island
- ElecDistr
- ElecGen
- ElecTrans
- Eleccome
- RegPrice
- Intercon
- Nuclear
- OCDE
- RatParis
- Government
- EU
- TempAverage
- PrepAverage
- Nsub
- Boperyear
- Bonorm
- ETIAverage
- ENSAverage

First, the variable *Country* is analysed. The frequency of appearance of every country in the dataset is shown in Figures X and X. All of the countries appear once per year, 14 times in total, with any exception.

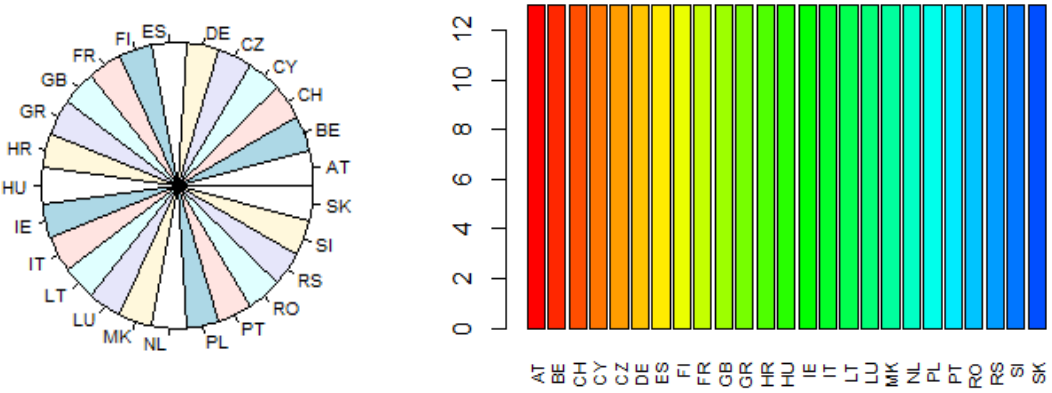


Figure 48

Economic variables:

- GDP per Capita:

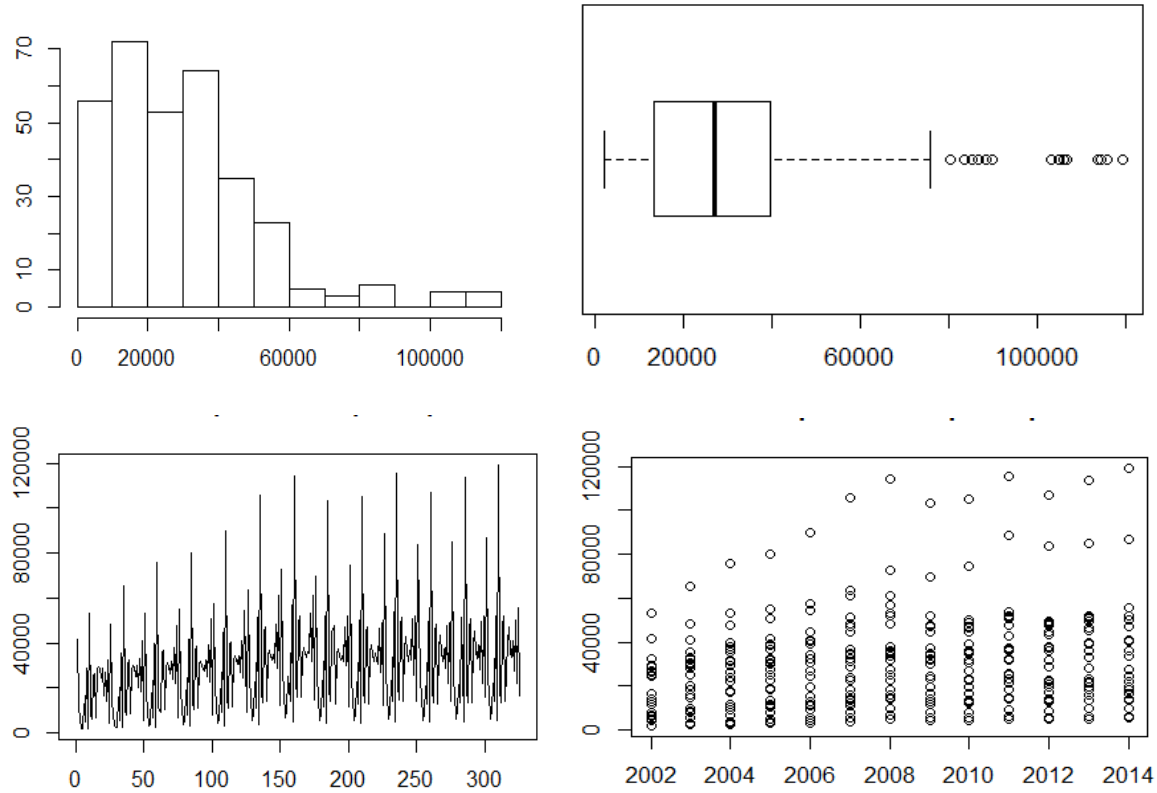


Figure 49

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1961   13467   27170   29883   39539  119225
## [1] "sd:  22018.5914185182"
## [1] "vc:  0.73683411718615"
```

- GINI Index

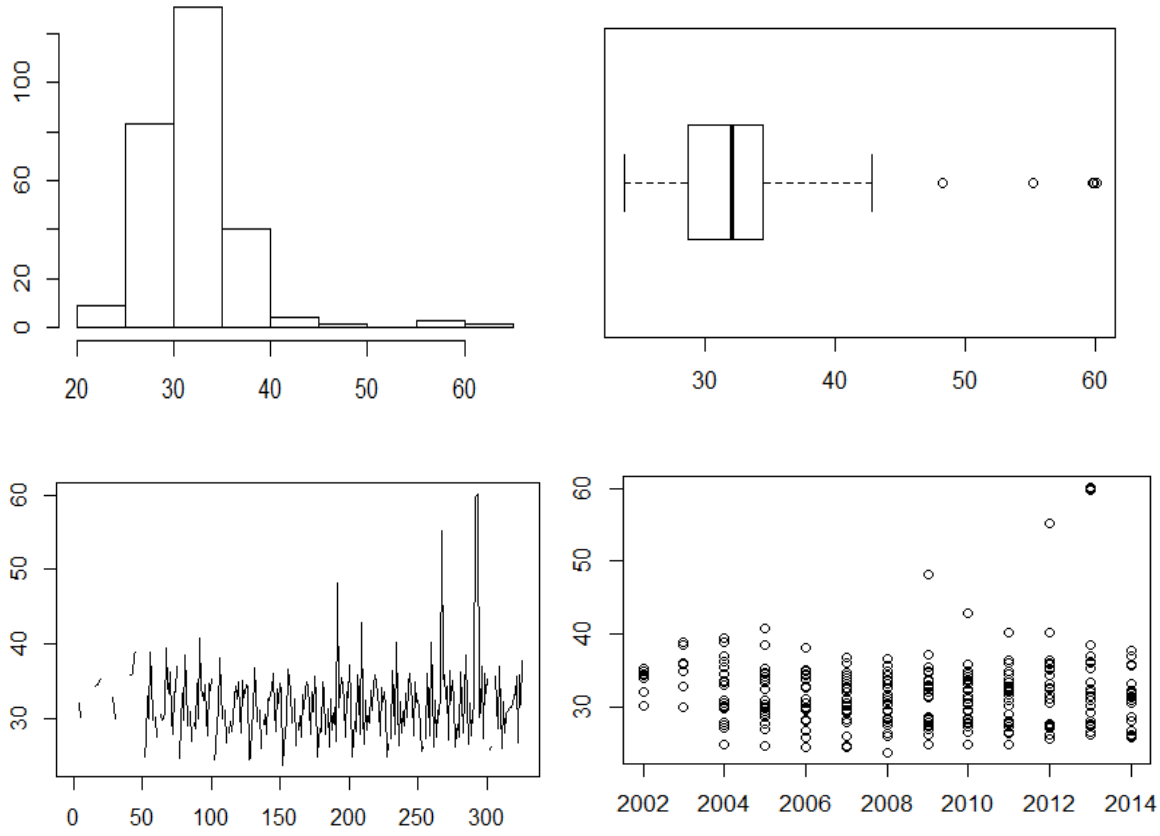


Figure 50

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   23.70   28.70   32.00   32.06   34.38   60.12      53
## [1] "sd:  4.94437111014013"
## [1] "vc:  0.154218716288817"
```

- Total Population:

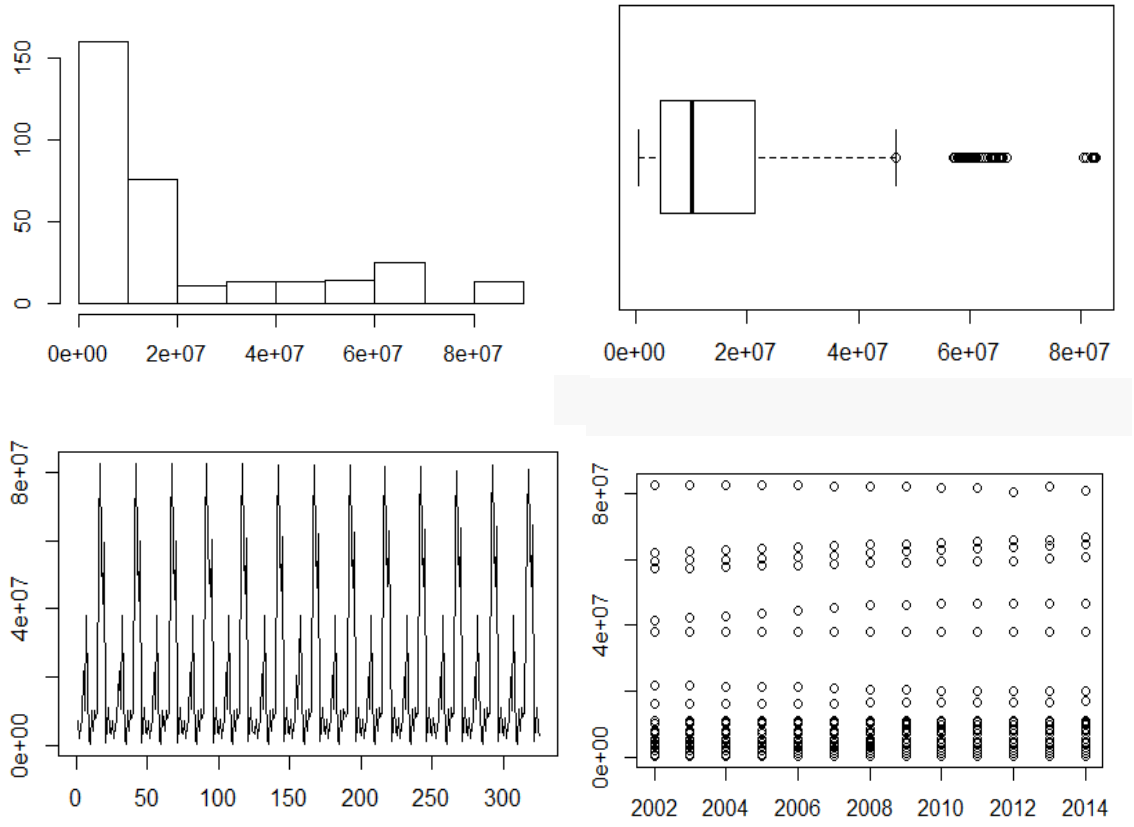
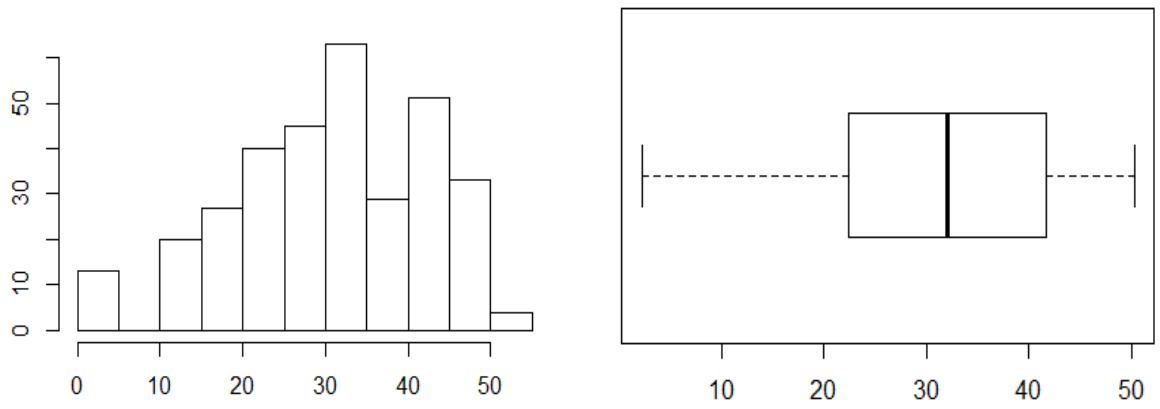


Figure 51

```
## [1] "Extended Summary Statistics"
##    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  446175  4440000 10038188 19659189 21319685 82534176
## [1] "sd:  23281265.7608497"
## [1] "vc:  1.18424345069122
```

Next two variables, *Rural and Urban Populations* are expressed here as compositional data and are measured in percentage of the *Total Population.*

- Rural Population:



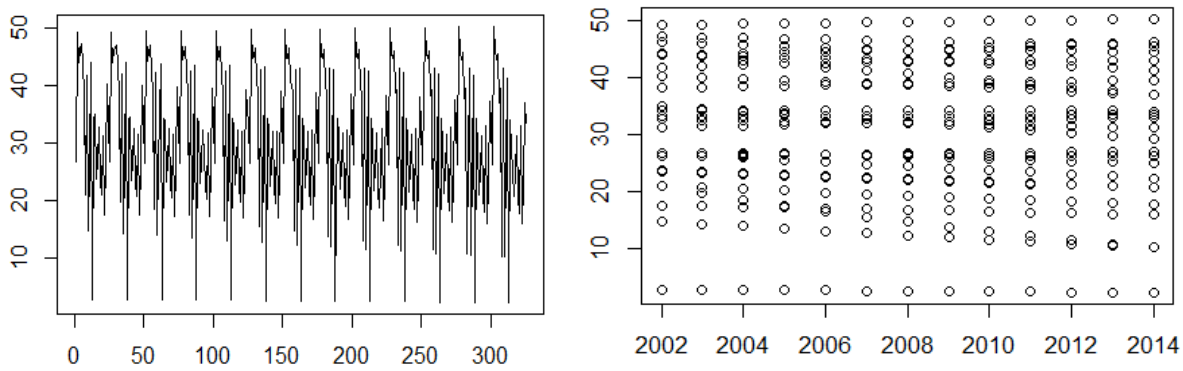


Figure 52

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.182  22.379  32.026  30.684  41.641  50.305
## [1] "sd:  11.9482403845515"
## [1] "vc:  0.389396479686303"
```

- Urban Population:

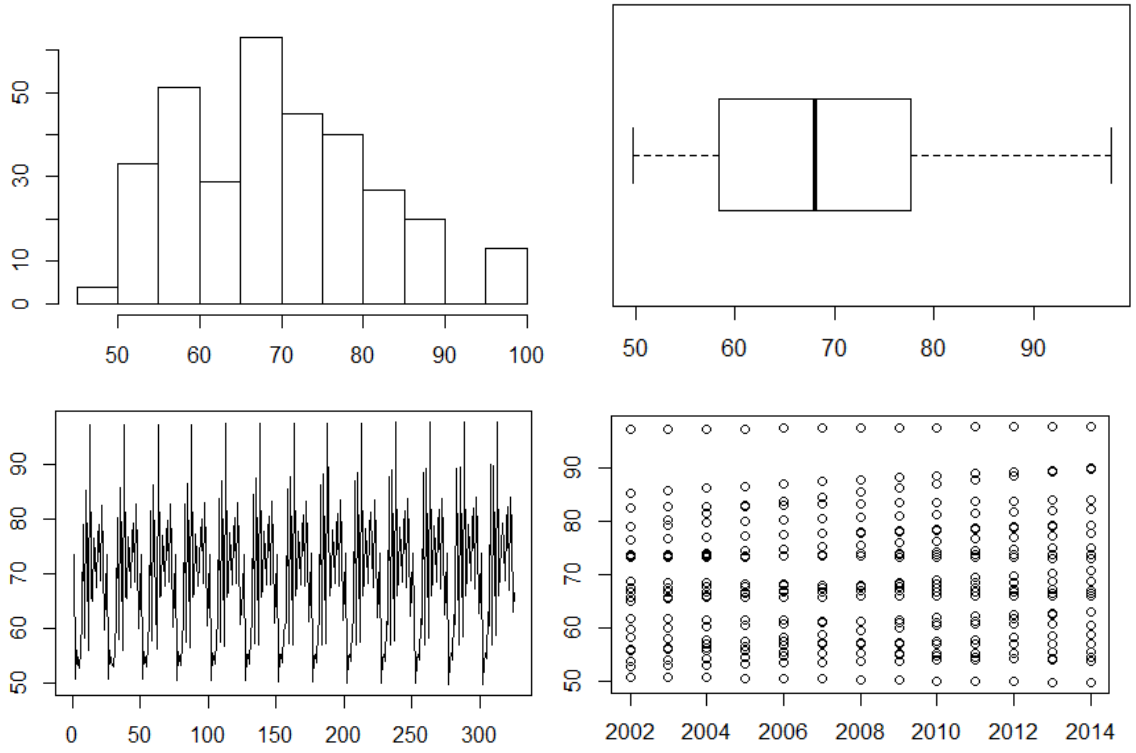


Figure 53

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   49.70   58.36   67.97   69.32   77.62   97.82
## [1] "sd:  11.9482403845515"
## [1] "vc:  0.17237347588106"
```

♣

- Corruption Ranking:

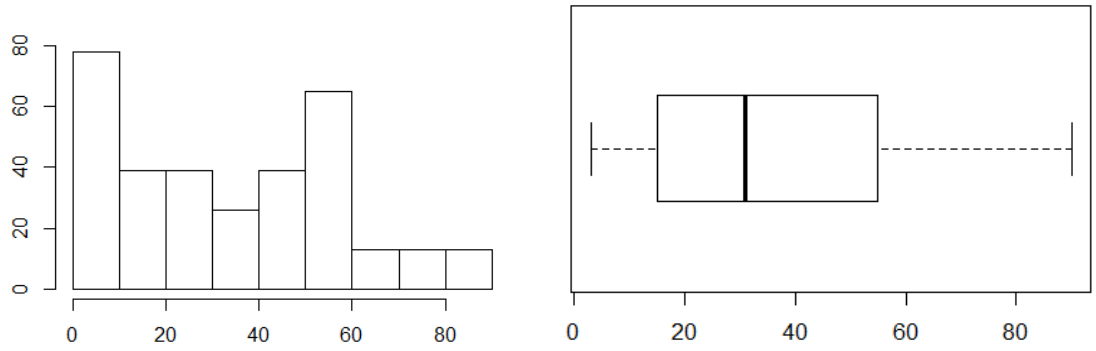


Figure 54

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00   15.00   31.00   35.84   55.00   90.00
## [1] "sd:  23.6663849748494"
## [1] "vc:  0.660334402200039"
```

- Democracy Ranking:

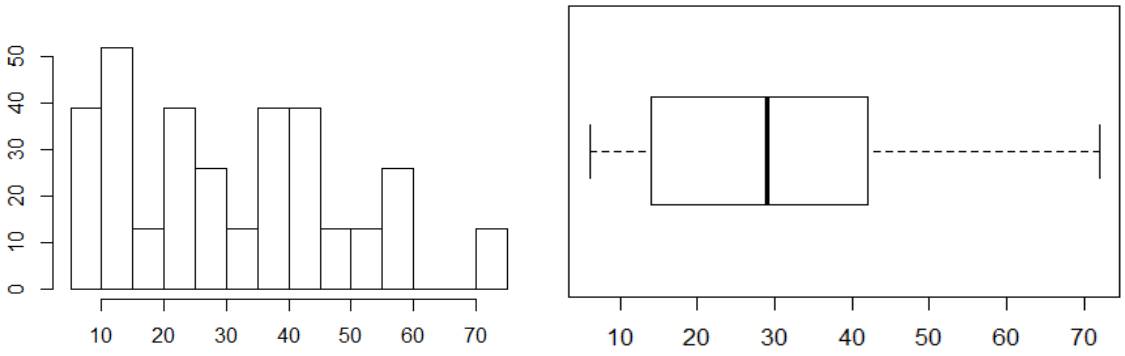


Figure 55

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##    6.00   14.00   29.00   31.08   42.00   72.00
## [1] "sd:  17.6155725158659"
## [1] "vc:  0.566781612479596"
```

- Belonging to the OCDE:

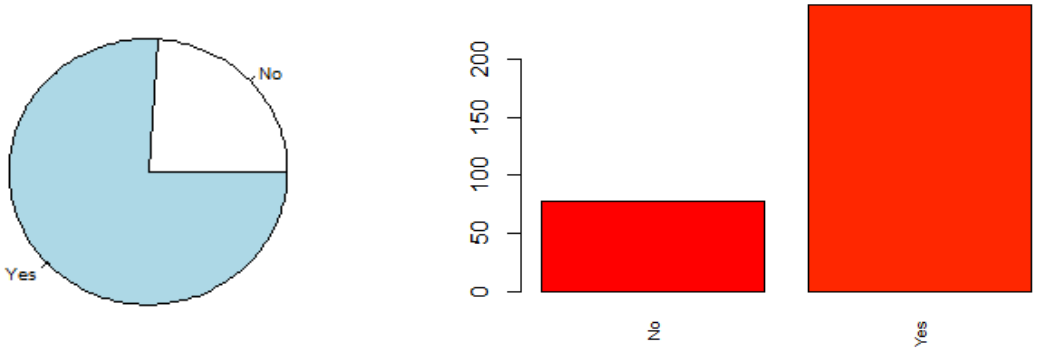


Figure 56

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
##  78 247
## [1] "Relative frequency table (proportions)"
##   No  Yes
## 0.24 0.76
## [1] "Frequency table sorted"
## Yes  No
## 247  78
## [1] "Relative frequency table (proportions) sorted"
##  Yes   No
## 0.76 0.24
```

- Type of Government:

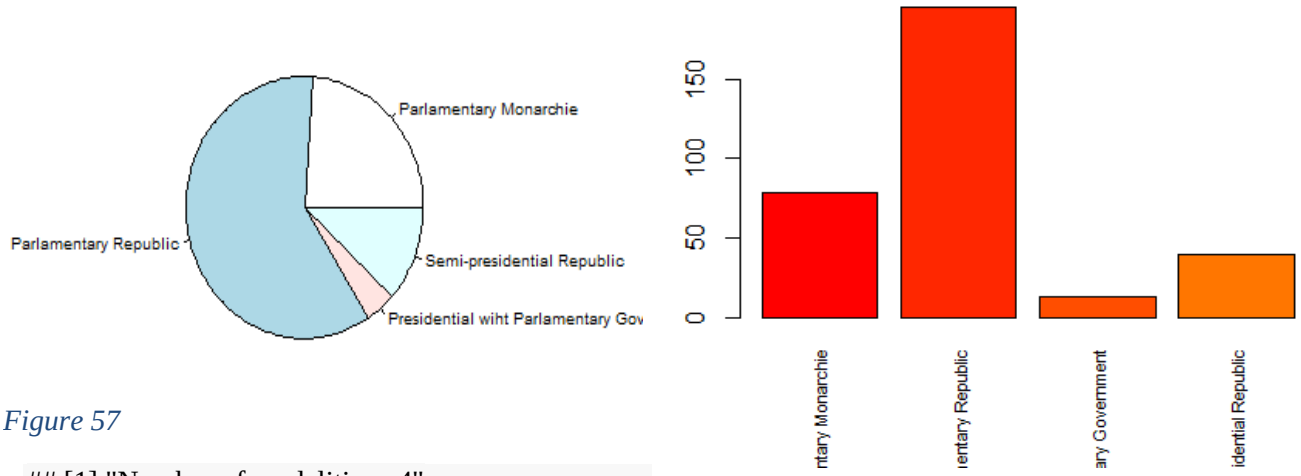


Figure 57

```
## [1] "Number of modalities:  4"
## [1] "Frequency table"
##                  Parlamentary Monarchie
##                                     78
##                   Parlamentary Republic
##                                    195
## Presidential wiht Parlamentary Government
##                                     13
##                Semi-presidential Republic
```

```
##                                39
## [1] "Relative frequency table (proportions)"
##                  Parlamentary Monarchie
##                                0.24
##                   Parlamentary Republic
##                                0.60
## Presidential wiht Parlamentary Government
##                                0.04
##              Semi-presidential Republic
##                                0.12
## [1] "Frequency table sorted"
##                   Parlamentary Republic
##                                 195
##                  Parlamentary Monarchie
##                                  78
##              Semi-presidential Republic
##                                  39
## Presidential wiht Parlamentary Government
##                                  13
## [1] "Relative frequency table (proportions) sorted"
##                   Parlamentary Republic
##                                0.60
##                  Parlamentary Monarchie
##                                0.24
##              Semi-presidential Republic
##                                0.12
## Presidential wiht Parlamentary Government
##                                0.04
```

- Belonging to the EU:

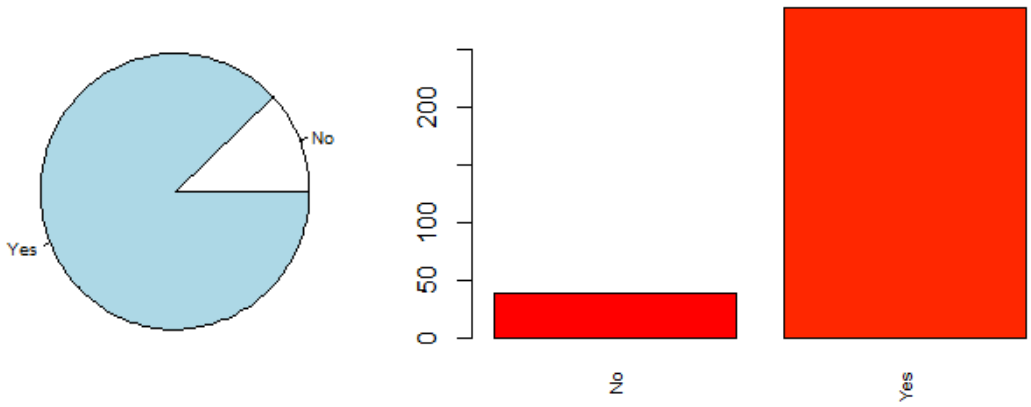


Figure 58

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
##  39 286
## [1] "Relative frequency table (proportions)"
##   No  Yes
## 0.12 0.88
## [1] "Frequency table sorted"
```

```
## Yes  No
## 286  39
## [1] "Relative frequency table (proportions) sorted"
##  Yes   No
## 0.88 0.12
```

Energy variables:

- Emissions of CO_2 ;

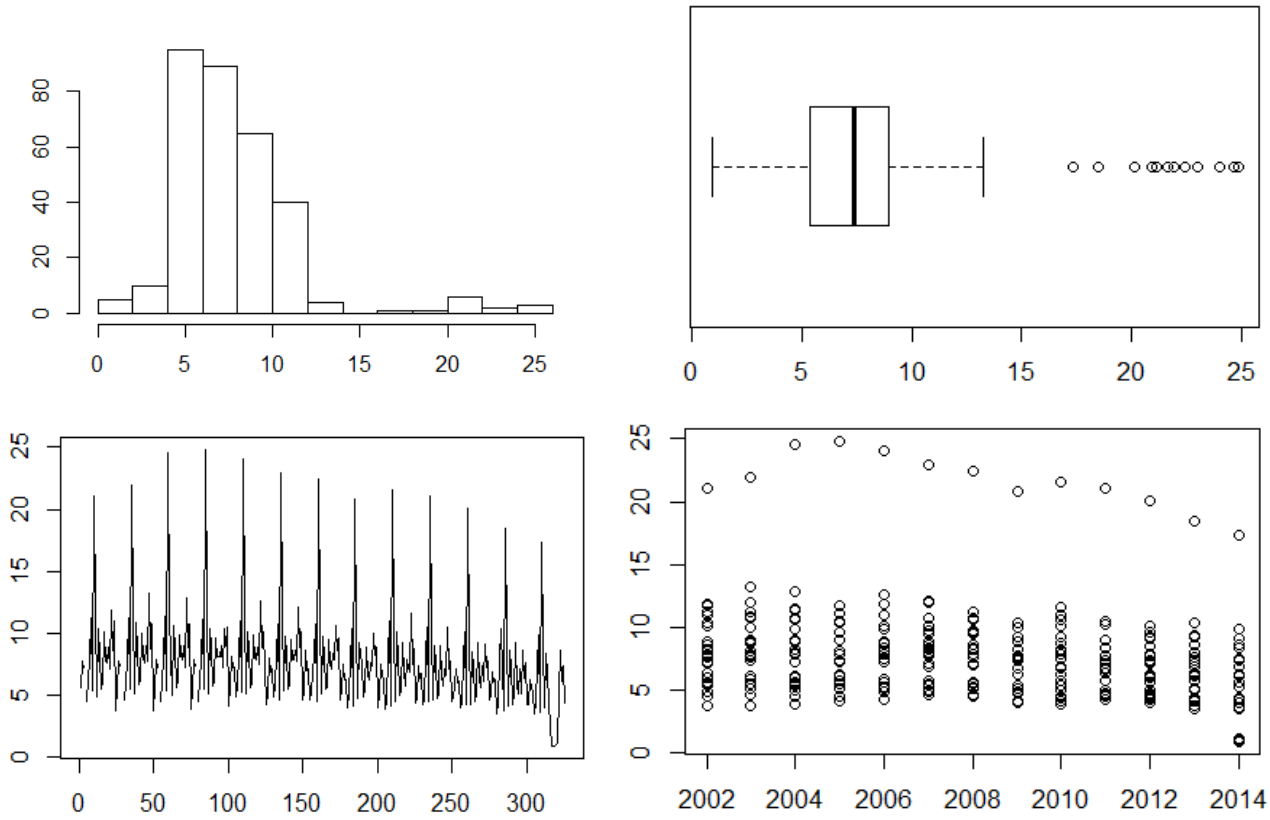


Figure 59

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.9197  5.3715  7.3506  7.7697  8.9829 24.8246       4
## [1] "sd:  3.69595727934577"
## [1] "vc:  0.475686363604303"
```

- Access to Electricity in terms of Total Population:

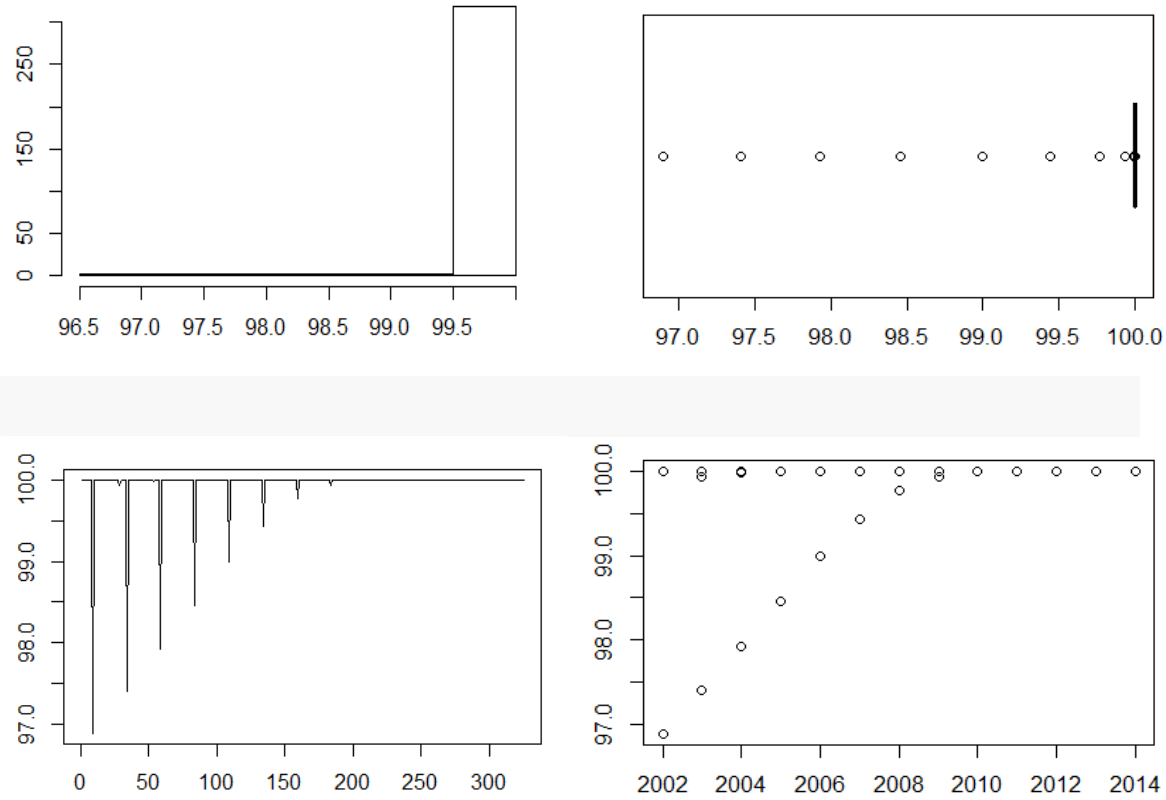


Figure 60

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##   96.89  100.00  100.00   99.97  100.00  100.00
## [1] "sd:  0.272407127448158"
## [1] "vc:  0.00272501443484283"
```

- Access to Electricity in terms of Rural Population:

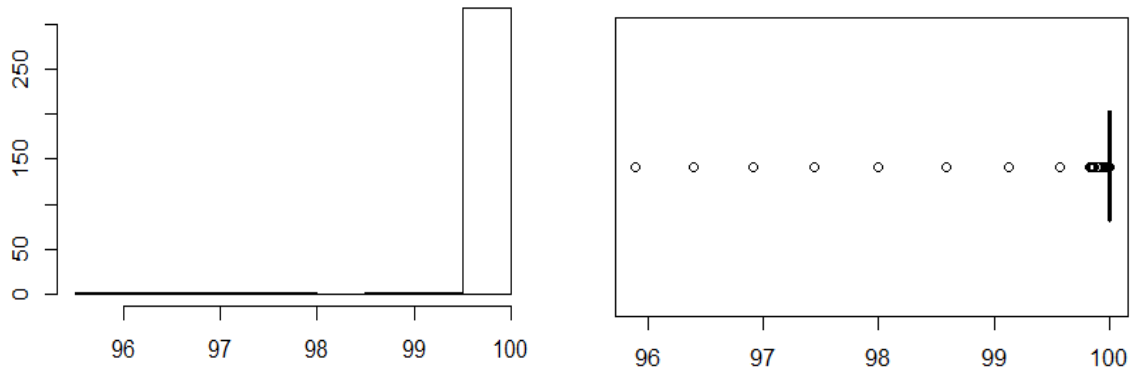
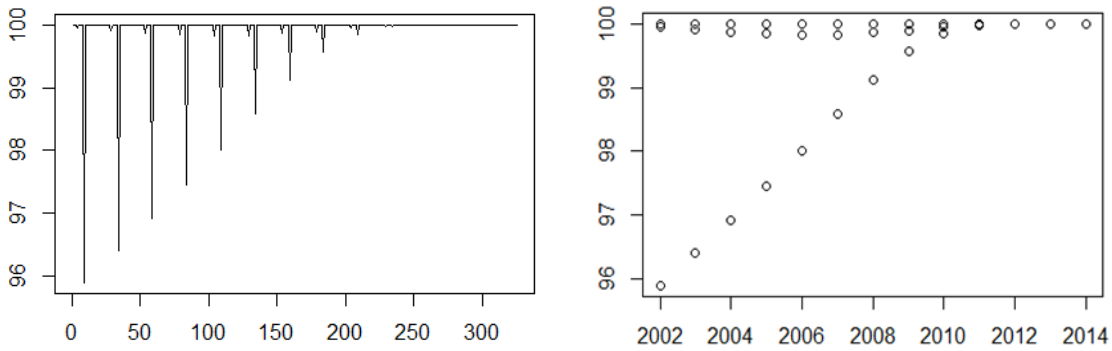


Figure 61

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   95.89  100.00  100.00   99.94  100.00  100.00
```

```
## [1] "sd:  0.400535721657227"
## [1] "vc:  0.00400774604429463"
```



Next four variables are compositional data. They express the percentage of electricity produced from different sources and, as a whole, they form the total amount of electricity produced every year in every country.

- Electricity Production from Hydrological Sources:

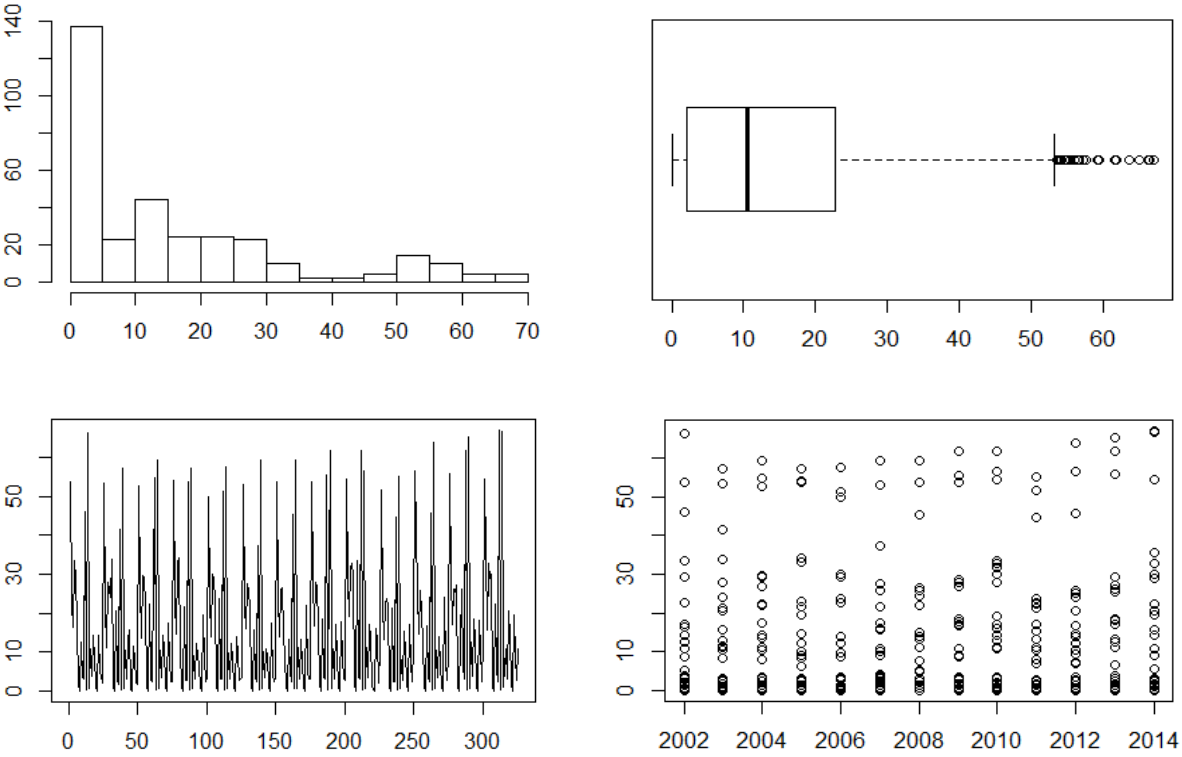


Figure 62

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.101  10.468  15.508  22.688  67.036
## [1] "sd:  17.505084446893"
## [1] "vc:  1.12875674358763"
```

- Electricity Produced from Nuclear Sources:

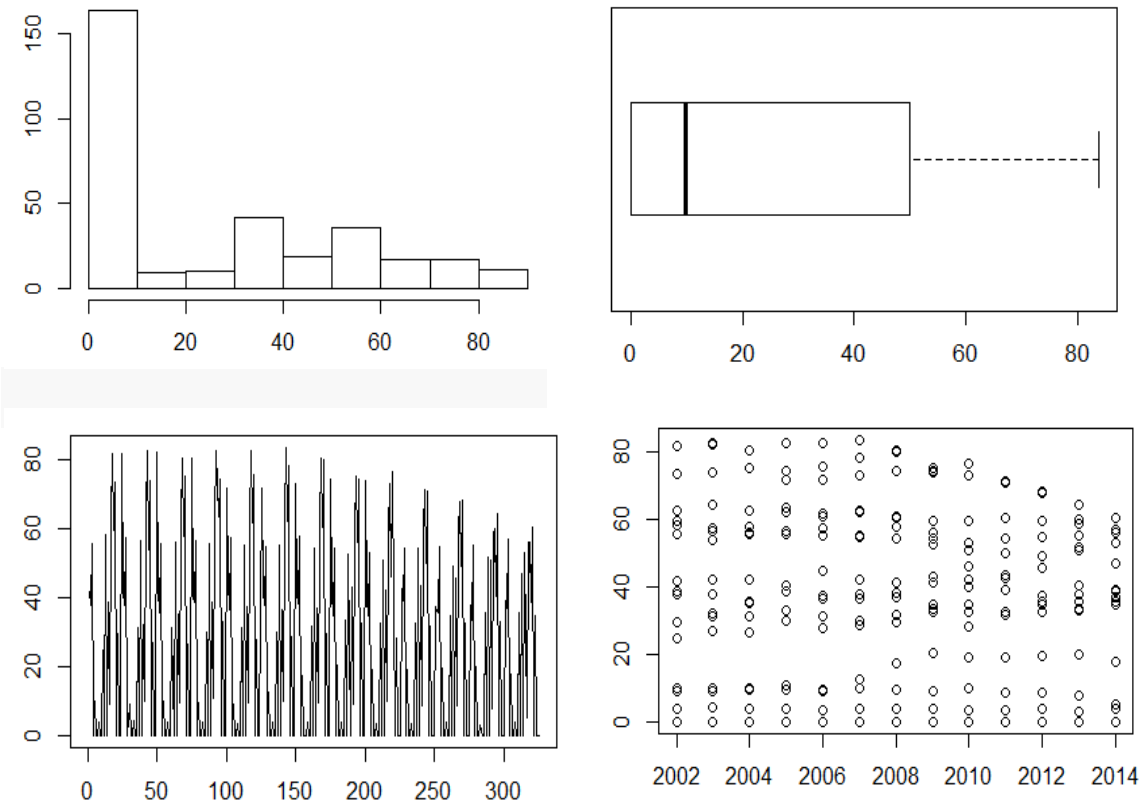


Figure 63

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   9.925  25.126  49.849  83.631
## [1] "sd:  27.3401702758991"
## [1] "vc:  1.08811395151089"
```

- Electricity Production from Oil, Gas and Coal Sources:

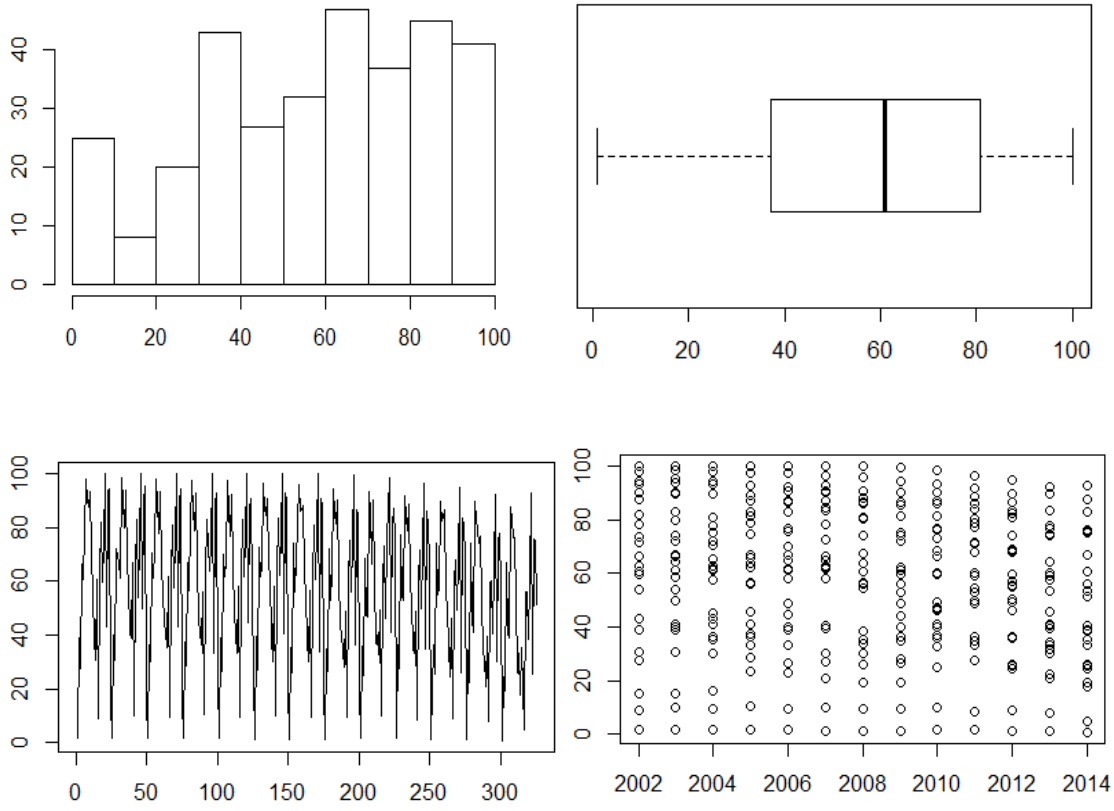
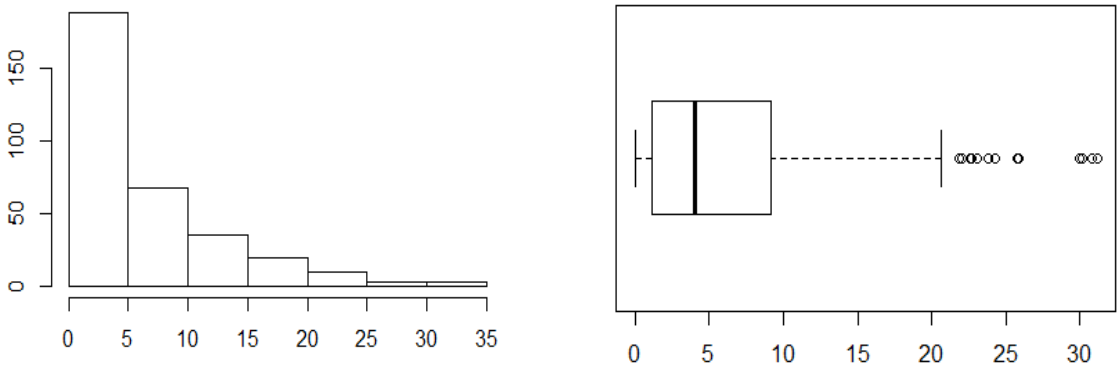


Figure 64

```
## [1] "Extended Summary Statistics"
##     Min.  1st Qu.   Median     Mean  3rd Q
u.     Max.
##   0.8031  37.1587  60.7521  57.8798  80.6594 100.0000
## [1] "sd:  26.9988601142554"
## [1] "vc:  0.466464485099953"
```

- Electricity Produced from Renewable Sources:



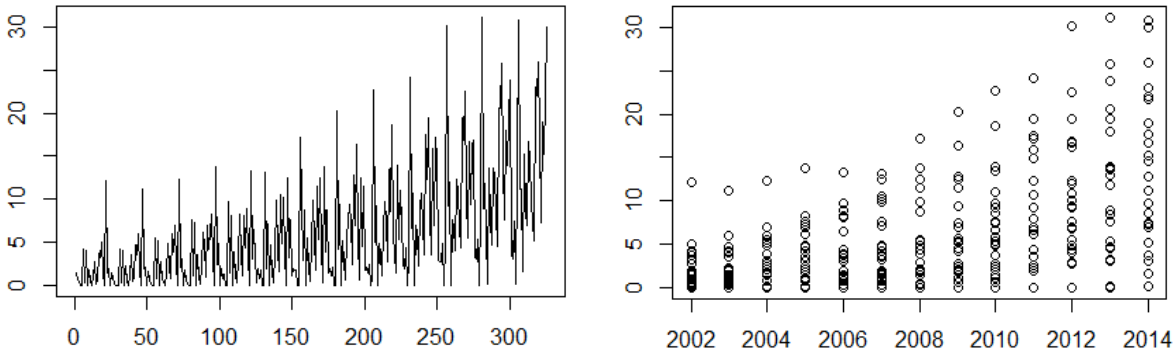


Figure 65

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.131   4.012   6.116   9.178  31.149
## [1] "sd:  6.55628133754136"
## [1] "vc:  1.07195487129935"
```

- Energy Imports:

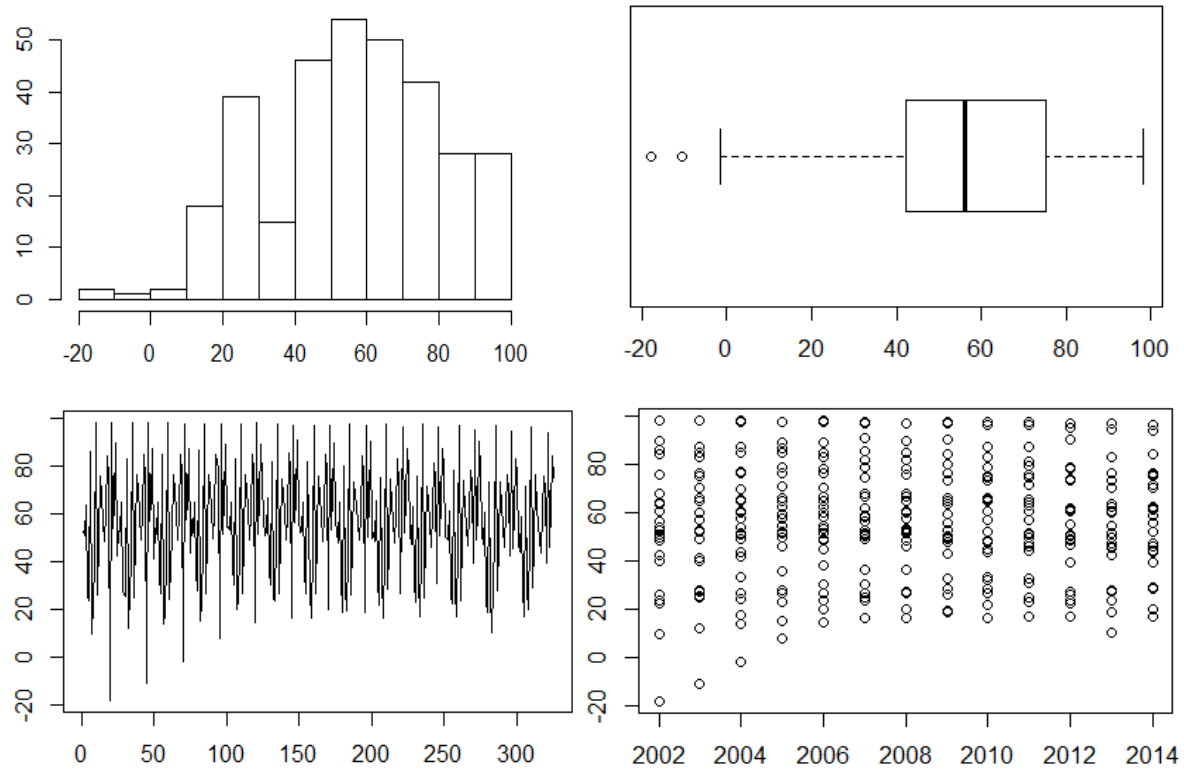


Figure 66

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -18.02   42.11   56.26   56.28   75.27   98.08
## [1] "sd:  23.8176362242681"
## [1] "vc:  0.42320521116045"
```

- Electric Power Consumed:

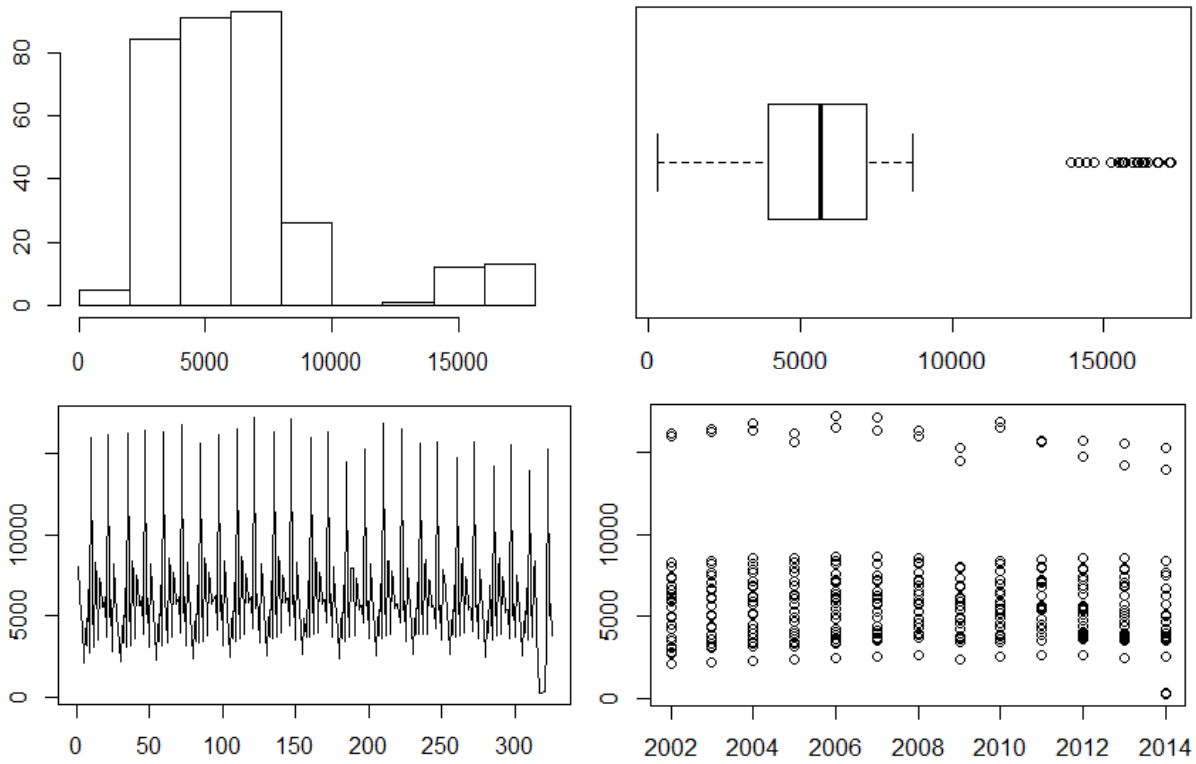


Figure 67

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##   263.6  3921.9  5661.6  6227.6  7187.8 17215.0
## [1] "sd:  3343.46879229361"
## [1] "vc:  0.536878213676627"
```

- Total Greenhouse Gas Emissions Change from 1990:

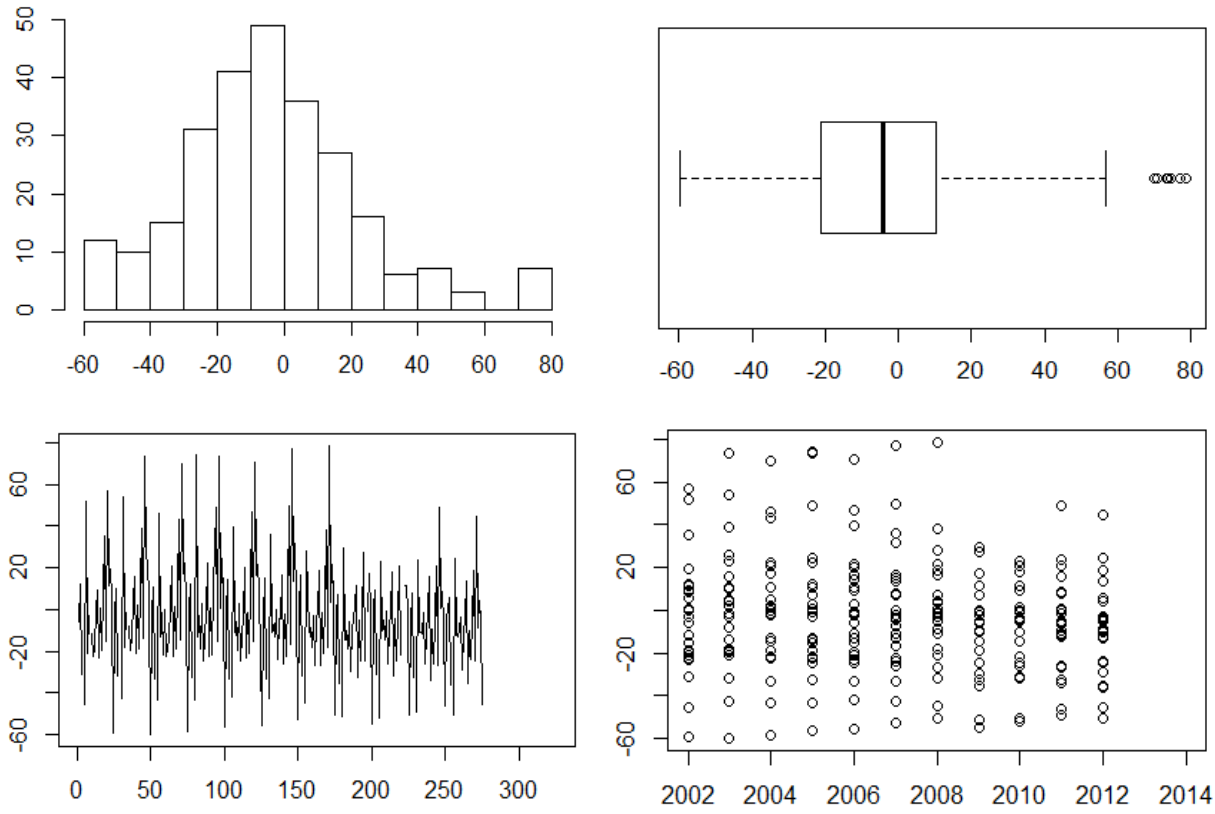
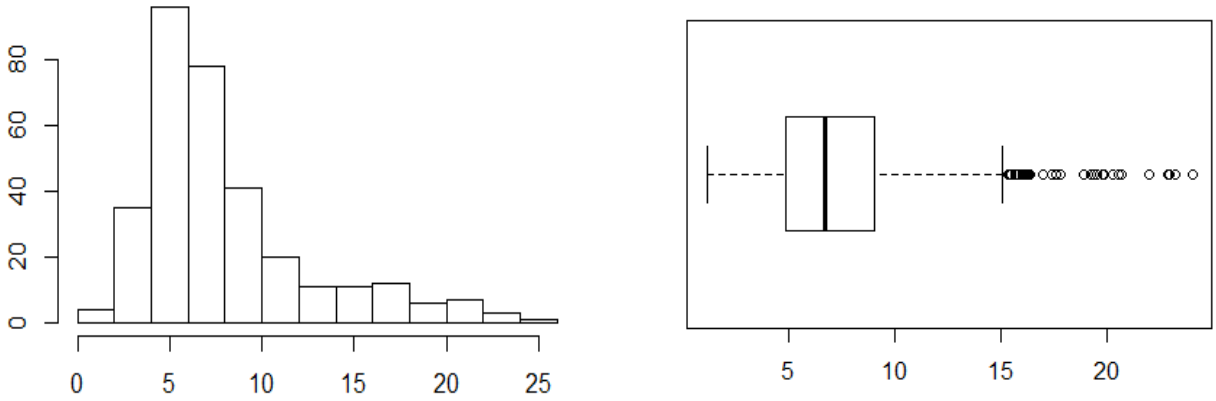


Figure 68

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.    NA's
## -59.905 -21.031  -4.257  -3.766  10.488  78.661     65
## [1] "sd:  26.6024018419098"
## [1] "vc:  -7.06316016451674"
```

- Electric Power Transmission and Distribution Losses:




```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.153   4.874   6.699   7.955   9.050  24.022
```

```
## [1] "sd:  4.53437620006654"
## [1] "vc:  0.570011647359259"
```

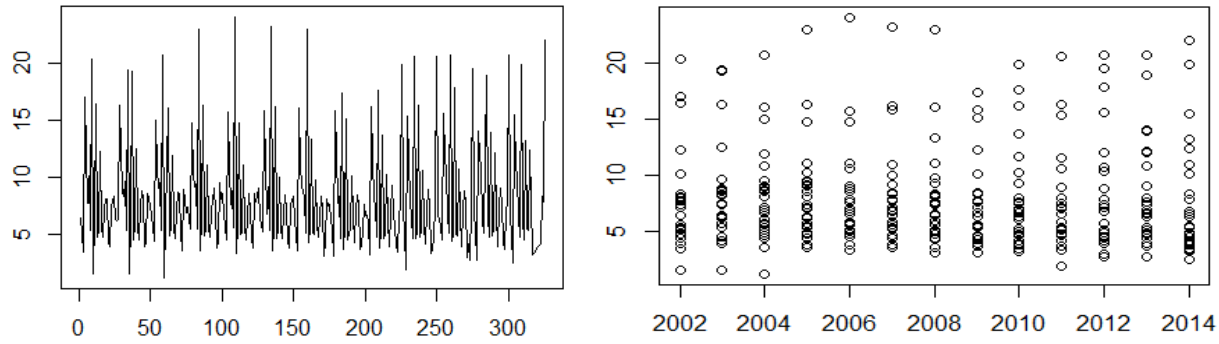


Figure 69

- Fuel Exports:

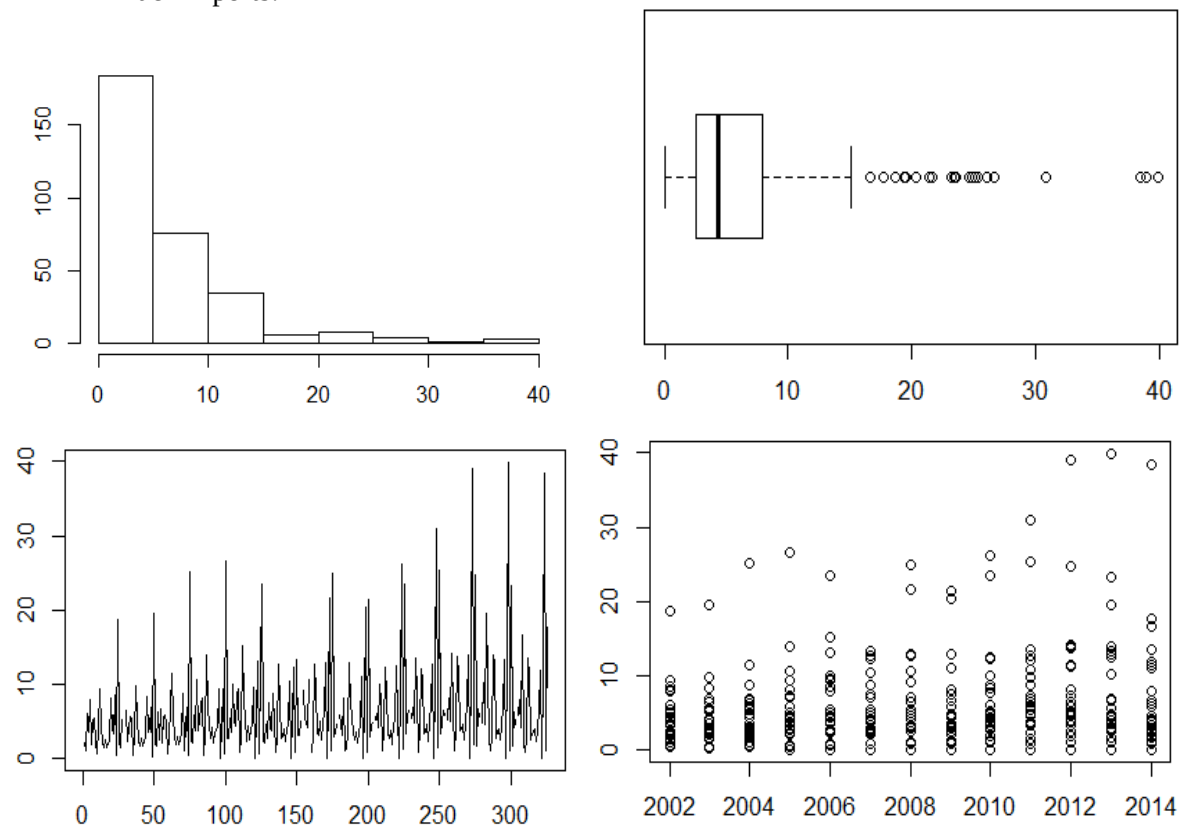


Figure 70

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   2.509   4.357   6.280   7.918  39.896       9
## [1] "sd:  6.31948568375844"
## [1] "vc:  1.00635240985677"
```

- Time required to get electricity:

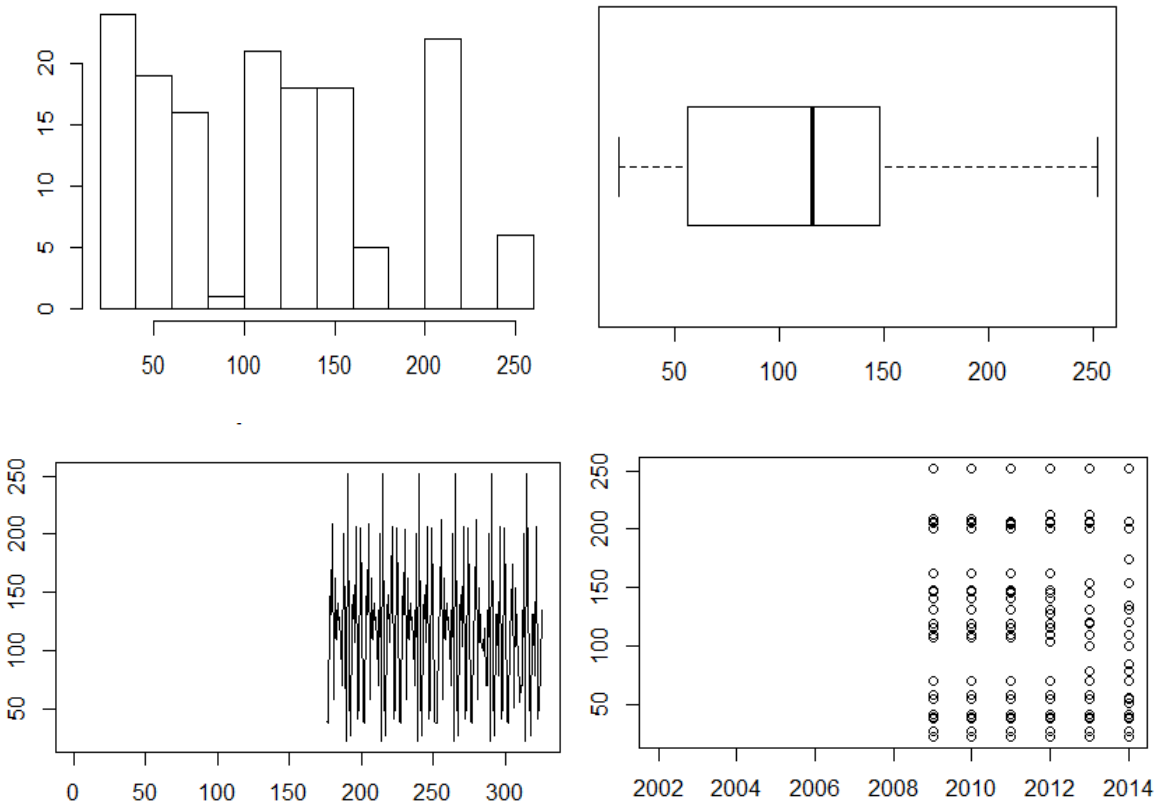
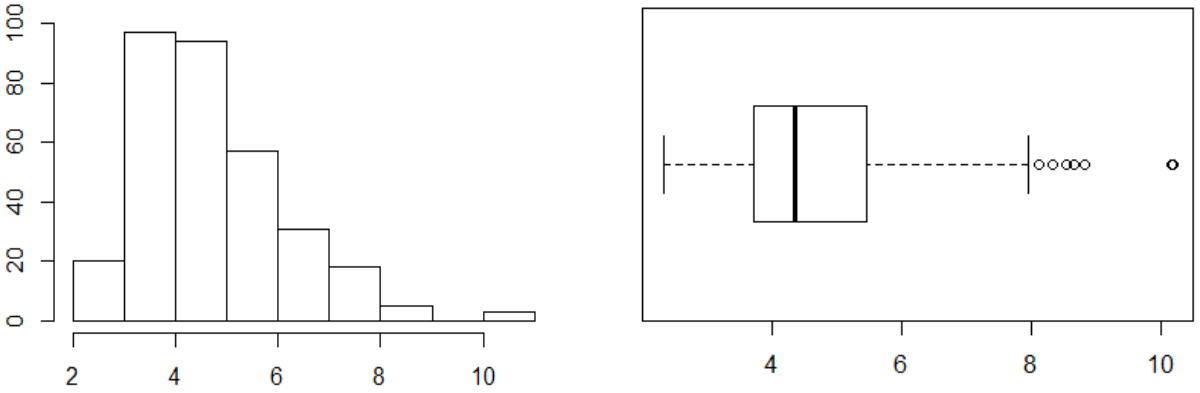


Figure 71

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   23.00   56.75  116.00  114.51  148.00  252.00     175
## [1] "sd:  63.6403207491597"
## [1] "vc:  0.555745945879604"
```

- Energy Intensity:



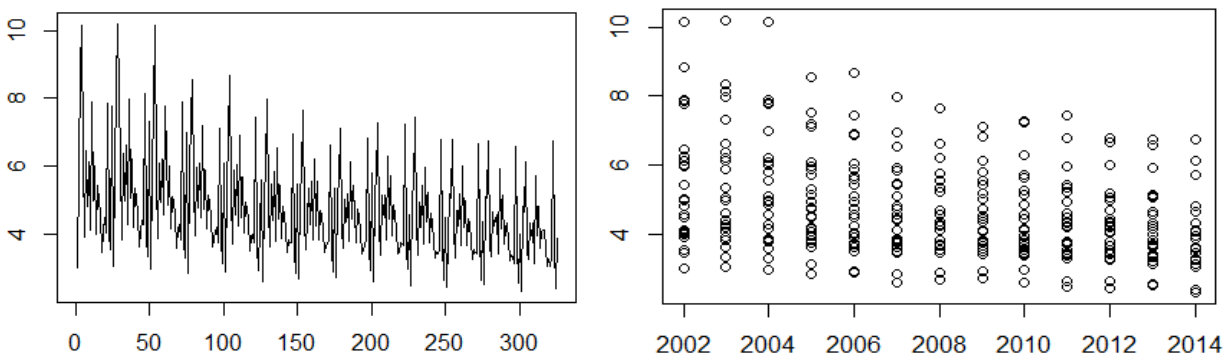


Figure 72

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##   2.318   3.726   4.358   4.731   5.466  10.194
## [1] "sd:  1.43535751676363"
## [1] "vc:  0.30337021295678"
```

- Electricity Distribution Market:

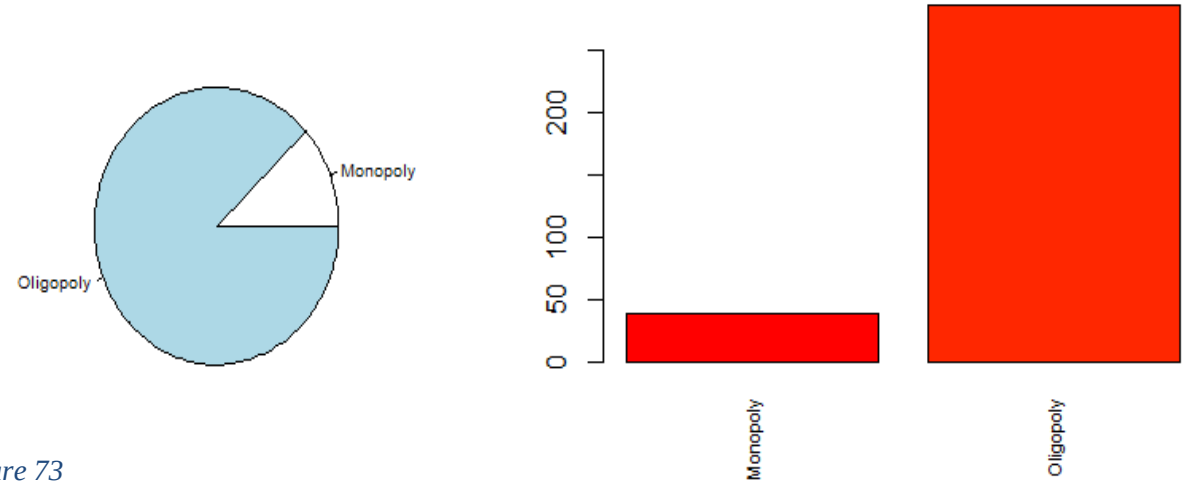


Figure 73

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  Monopoly Oligopoly
##        39       286
## [1] "Relative frequency table (proportions)"
##  Monopoly Oligopoly
##      0.12      0.88
## [1] "Frequency table sorted"
## Oligopoly  Monopoly
##       286        39
## [1] "Relative frequency table (proportions) sorted"
## Oligopoly  Monopoly
##      0.88      0.12
```

- Electricity Generation Market:

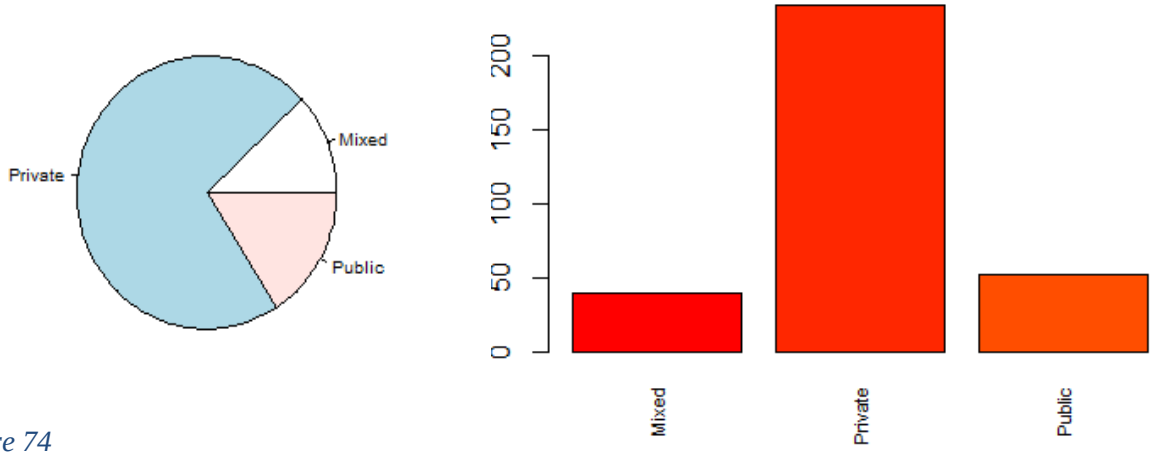


Figure 74

```
## [1] "Number of modalities:  3"
## [1] "Frequency table"
##   Mixed Private  Public
##      39     234      52
## [1] "Relative frequency table (proportions)"
##   Mixed Private  Public
##    0.12    0.72    0.16
## [1] "Frequency table sorted"
## Private  Public   Mixed
##     234      52      39
## [1] "Relative frequency table (proportions) sorted"
## Private  Public   Mixed
##    0.72    0.16    0.12
```

- Electricity Transmission Market:

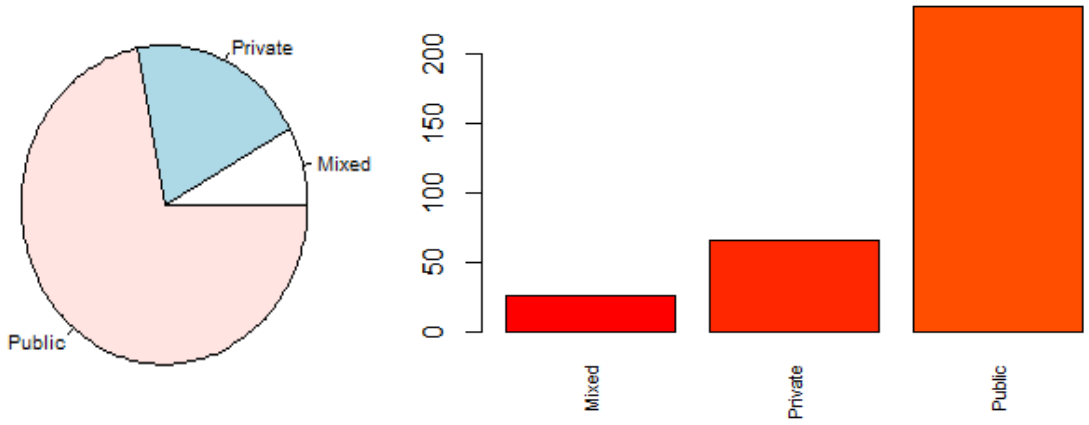


Figure 75

```
## [1] "Number of modalities:  3"
## [1] "Frequency table"
##   Mixed Private  Public
##      26      65     234
```

```
## [1] "Relative frequency table (proportions)"
##   Mixed Private  Public
##    0.08    0.20    0.72
## [1] "Frequency table sorted"
##  Public Private   Mixed
##     234      65      26
## [1] "Relative frequency table (proportions) sorted"
##  Public Private   Mixed
##    0.72    0.20    0.08
```

- Electricity Commercialization Market:

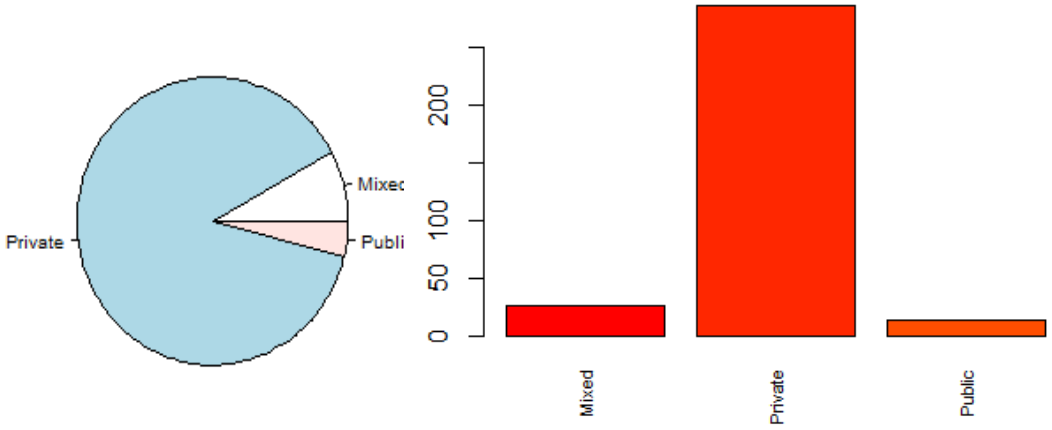


Figure 76

```
## [1] "Number of modalities:  3"
## [1] "Frequency table"
##   Mixed Private  Public
##      26     286      13
## [1] "Relative frequency table (proportions)"
##   Mixed Private  Public
##    0.08    0.88    0.04
## [1] "Frequency table sorted"
## Private   Mixed  Public
##     286      26      13
## [1] "Relative frequency table (proportions) sorted"
## Private   Mixed  Public
##    0.88    0.08    0.04
```

- Regulated Electricity Prices:

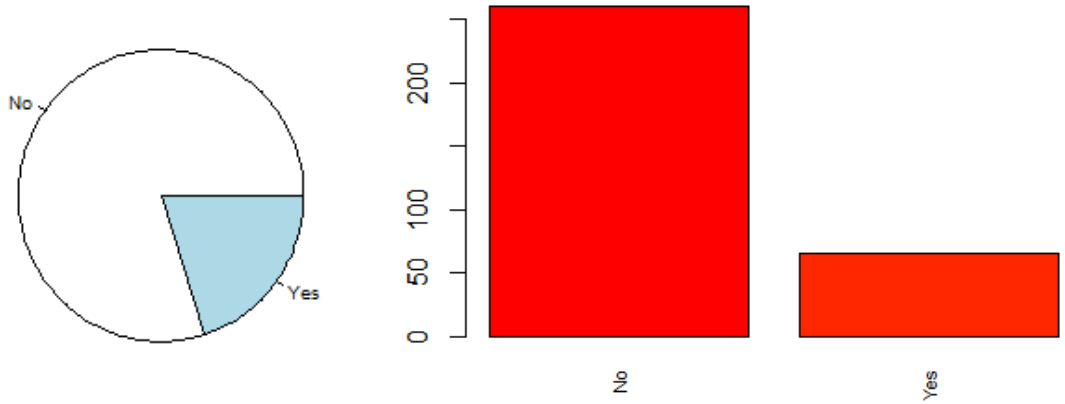


Figure 77

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
## 260  65
## [1] "Relative frequency table (proportions)"
##  No Yes
## 0.8 0.2
## [1] "Frequency table sorted"
##  No Yes
## 260  65
## [1] "Relative frequency table (proportions) sorted"
##  No Yes
## 0.8 0.2
```

- Interconnected Electric System:

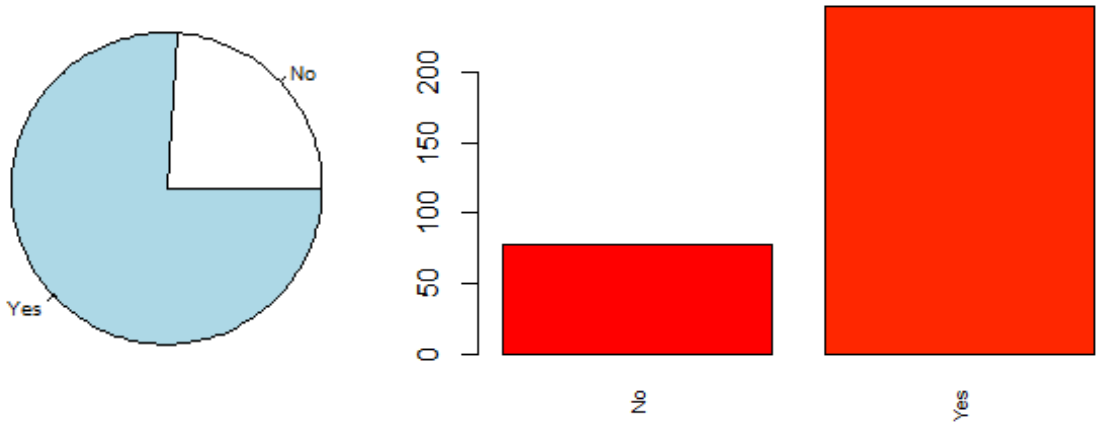


Figure 78

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
##  78 247
## [1] "Relative frequency table (proportions)"
##   No  Yes
## 0.24 0.76
## [1] "Frequency table sorted"
```

```
## Yes  No
## 247  78
## [1] "Relative frequency table (proportions) sorted"
##  Yes   No
## 0.76 0.24
```

- Nuclear Plants

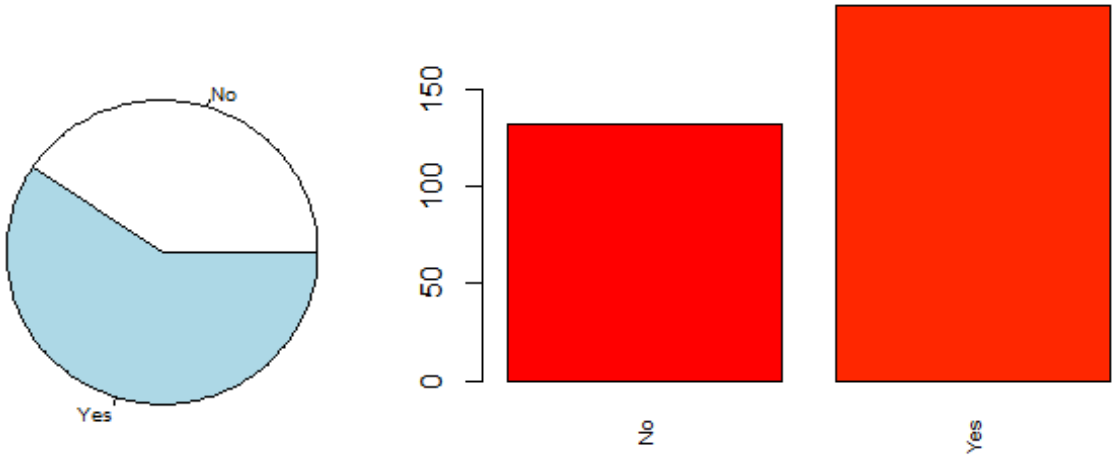
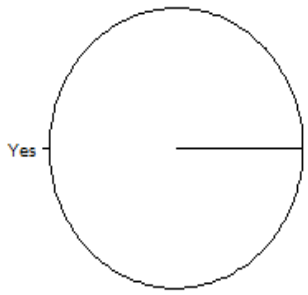


Figure 79

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
## 132 193
## [1] "Relative frequency table (proportions)"
##        No       Yes
## 0.4061538 0.5938462
## [1] "Frequency table sorted"
## Yes  No
## 193 132
## [1] "Relative frequency table (proportions) sorted"
##       Yes        No
## 0.5938462 0.4061538
```

- Ratification of the Paris Agreement:



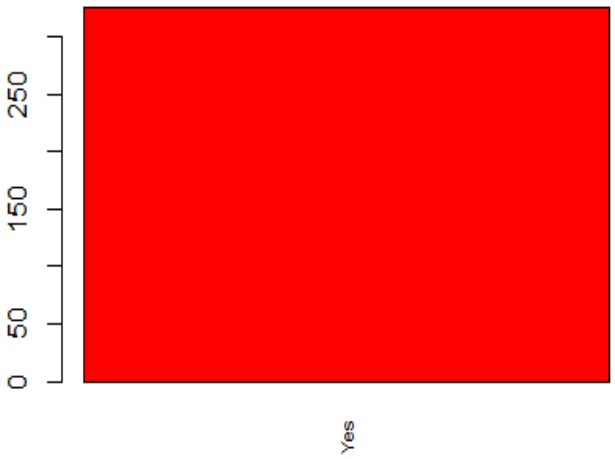
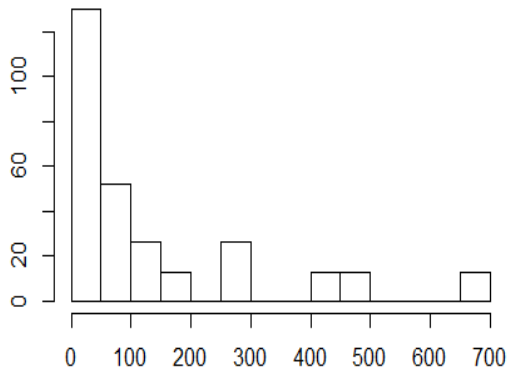


Figure 80

```
## [1] "Number of modalities:  1"
## [1] "Frequency table"
## Yes
## 325
## [1] "Relative frequency table (proportions)"
## Yes
##   1
## [1] "Frequency table sorted"
## Yes
## 325
## [1] "Relative frequency table (proportions) sorted"
## Yes
##   1
```

- Number of electric substations in the country:



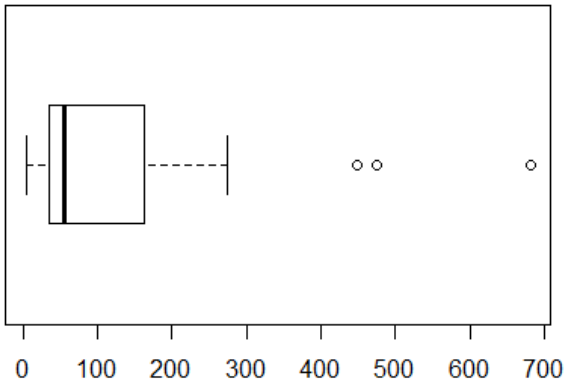


Figure 81

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Ma
x.    NA's
```

```
##    4.0   36.0   55.5  141.9  162.0  681.0     39
## [1] "sd:  176.269248612399"
## [1] "vc:  1.24252594343889"
```

- Number of Blackouts per year:

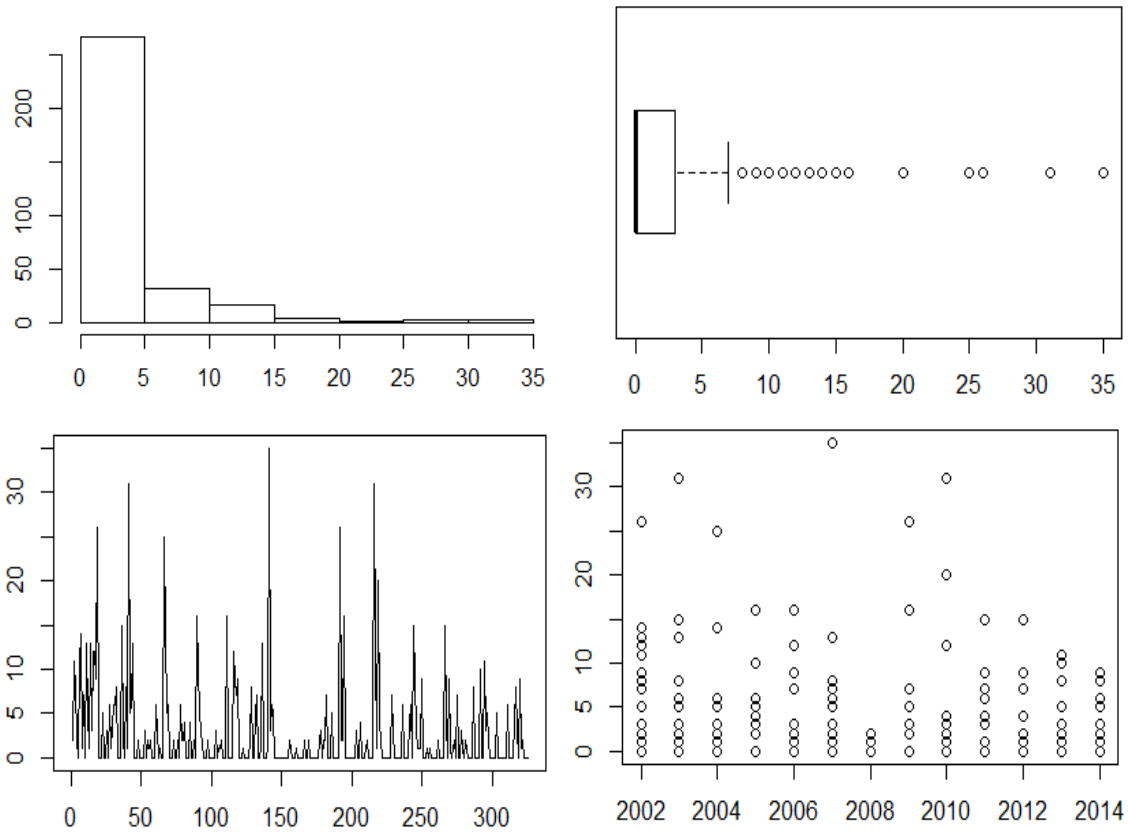


Figure 82

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##   0.000   0.000   0.000   2.708   3.000  35.000
## [1] "sd:  5.2795993257856"
## [1] "vc:  1.94985202372763"
```

- Number of Blackouts per year normalized by the number of electric substations in the country:

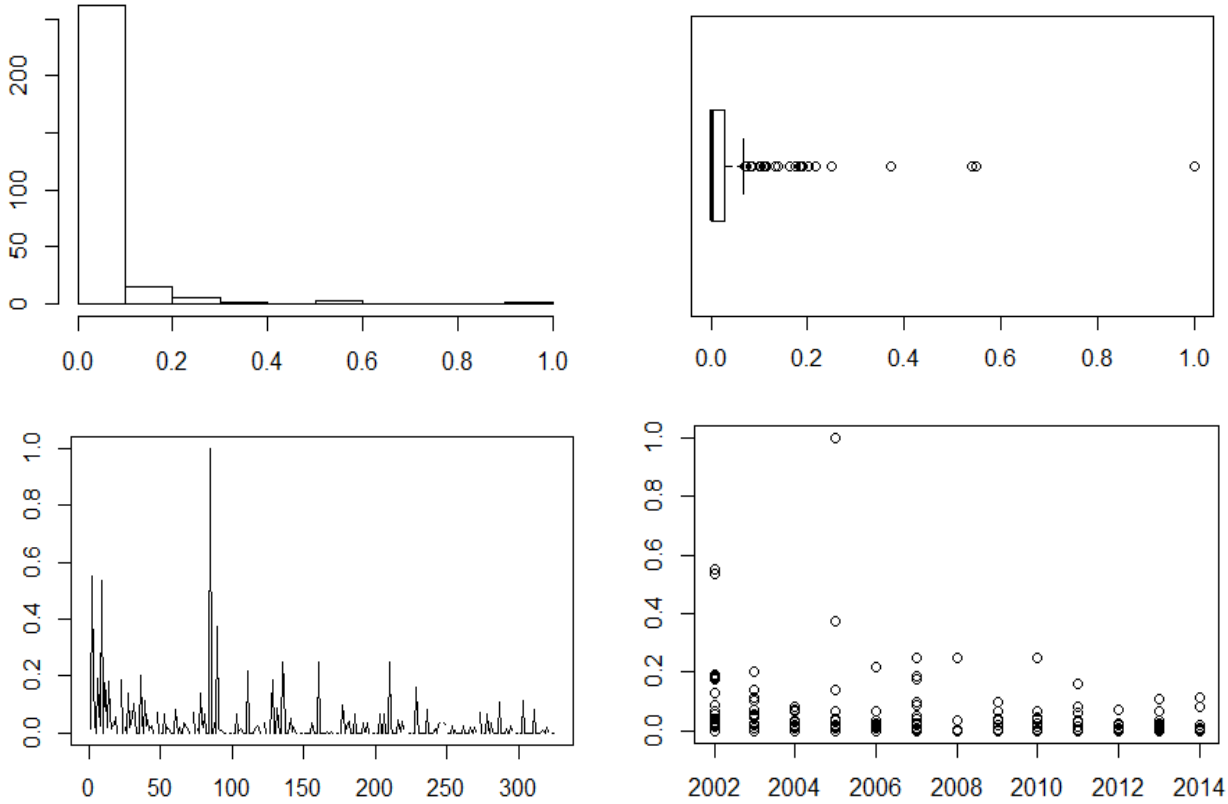


Figure 83

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.00000 0.00000 0.00000 0.03303 0.02733 1.00000      39
## [1] "sd:  0.088669262183974"
## [1] "vc:  2.68482457885605"
```

- Average Energy Not Supplied during the Failure:

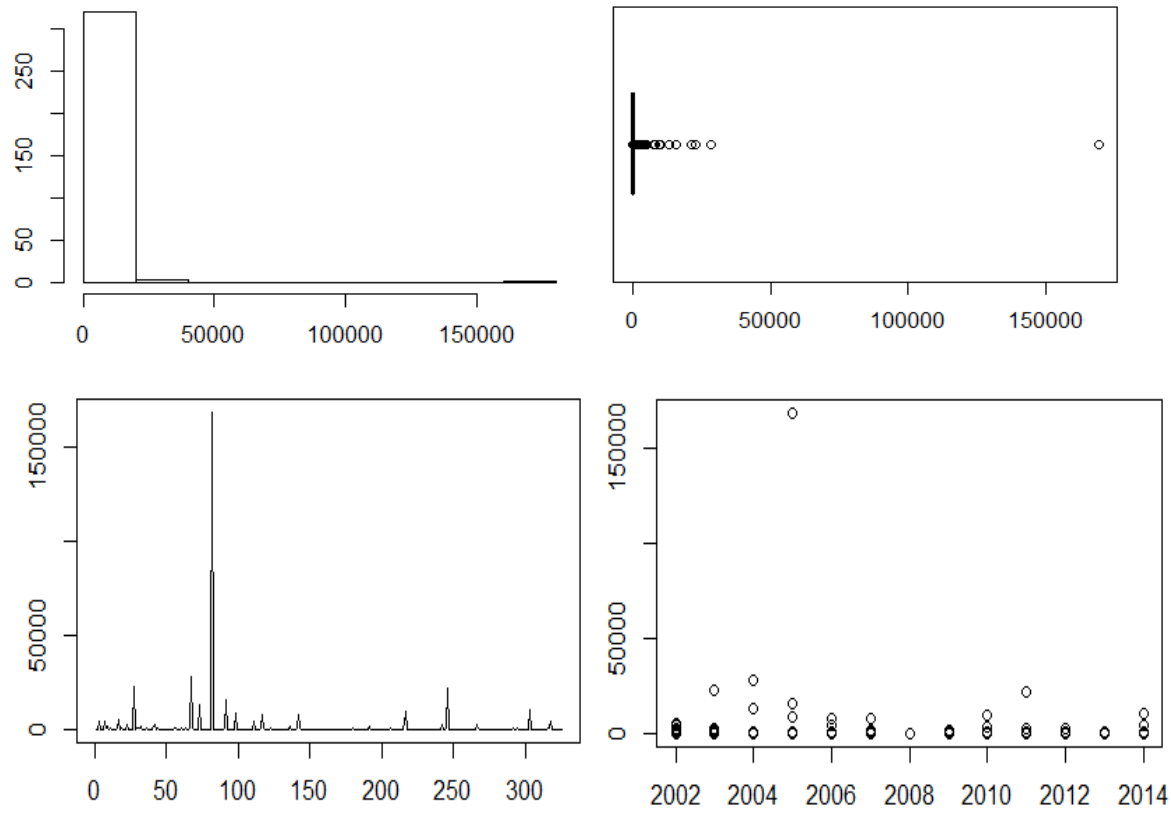


Figure 84

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       0       0    1177      75  168933
## [1] "sd:  9763.77593128956"
## [1] "vc:  8.29304844452864"
```

- Average Equivalent Time of the Blackout Taking into Account the Amount of Service during 12 Months:

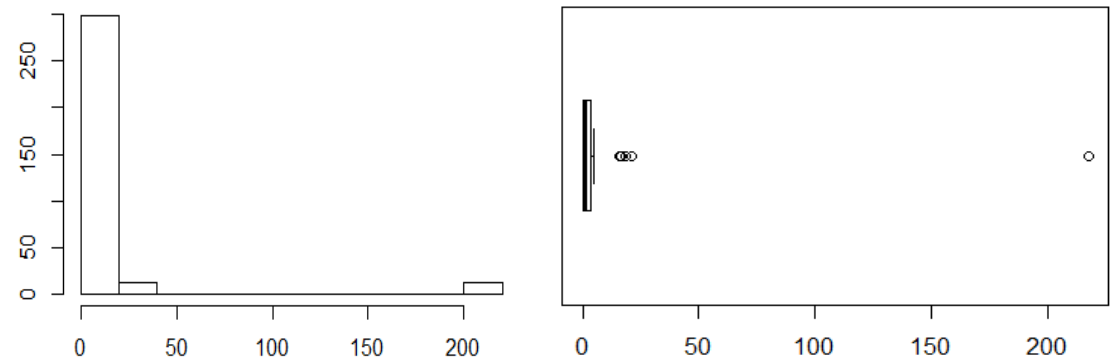


Figure 85

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.12    0.91   12.41    3.45  217.72
```

```
## [1] "sd:  42.4486404404833"
## [1] "vc:  3.41921851161229"
```

Climate variables:

- Type of Climate:

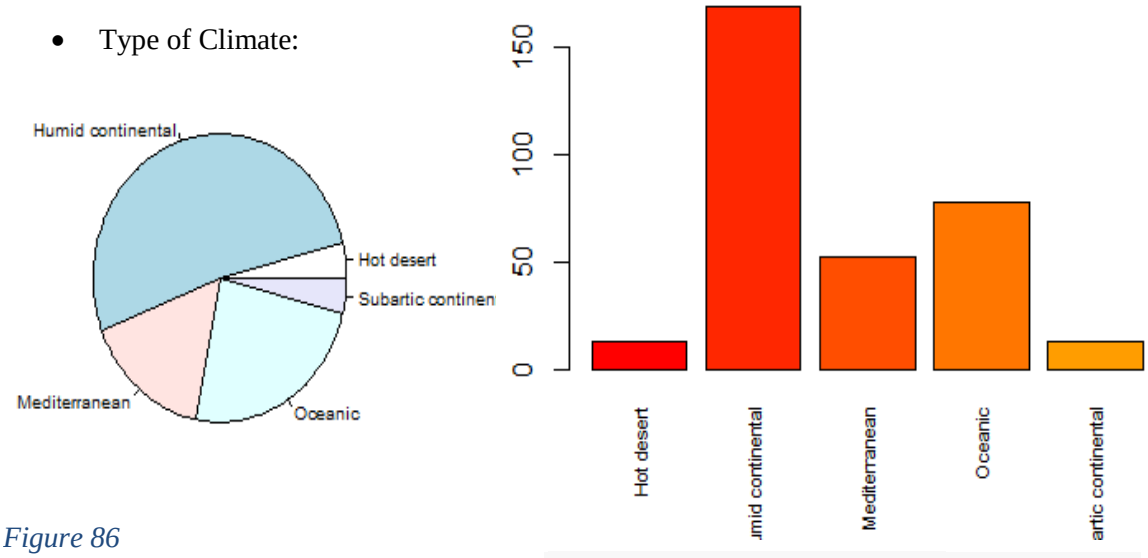


Figure 86

```
## [1] "Number of modalities:  5"
## [1] "Frequency table"
##          Hot desert    Humid continental       Mediterranean
##                  13                  169                  52
##            Oceanic Subartic continental
##                  78                   13
## [1] "Relative frequency table (proportions)"
##          Hot desert    Humid continental       Mediterranean
##                0.04                0.52                0.16
##            Oceanic Subartic continental
##                0.24                0.04
## [1] "Frequency table sorted"
##
##    Humid continental            Oceanic       Mediterranean
##                  169                 78                  52
##          Hot desert Subartic continental
##                   13                  13
## [1] "Relative frequency table (proportions) sorted"
##    Humid continental            Oceanic       Mediterranean
##                 0.52                0.24                0.16
##          Hot desert Subartic continental
##                 0.04                0.04
```

- Island:

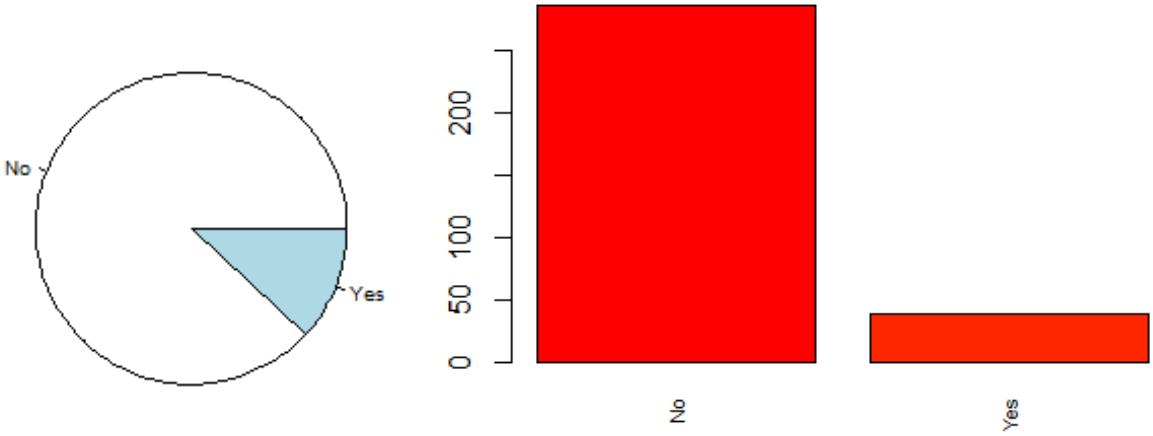


Figure 87

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
## 286  39
## [1] "Relative frequency table (proportions)"
##   No  Yes
## 0.88 0.12
## [1] "Frequency table sorted"
##  No Yes
## 286  39
## [1] "Relative frequency table (proportions) sorted"
##   No  Yes
## 0.88 0.12
```

- Average Temperature:

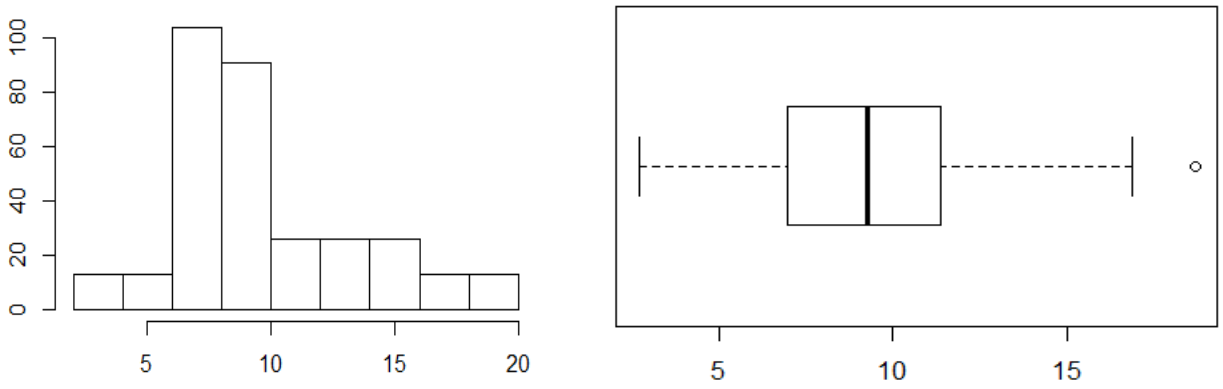


Figure 88

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.700   7.000   9.300   9.852  11.400  18.700
## [1] "sd:  3.76474516705175"
## [1] "vc:  0.382130041316662"
```

- Average Precipitation:

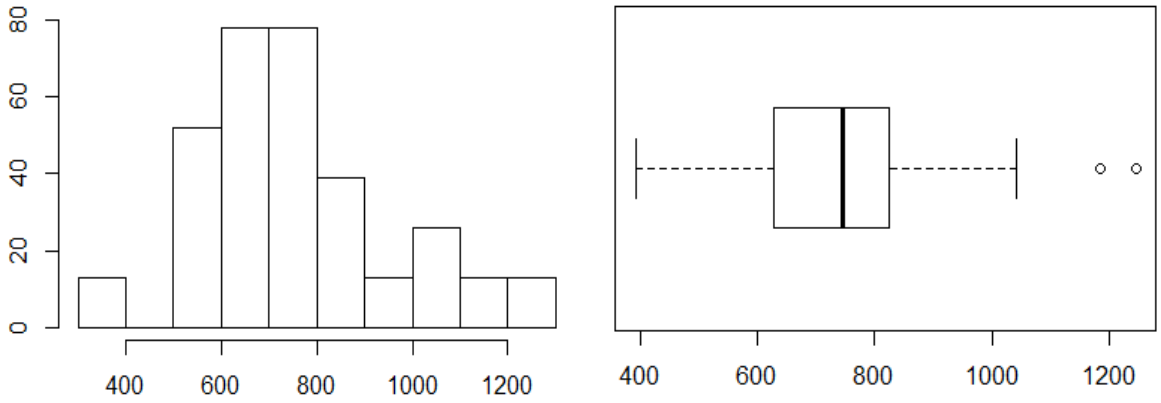


Figure 89

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##   391.1   626.2   745.6   761.7   825.4  1245.8
## [1] "sd:  202.01949226941"
## [1] "vc:  0.265226040412098"
```

8.2 Appendix B. Basic descriptive analysis for the major events dataset

The division of the variables into groups depending on their nature is the one stated in section 4.4.

The basic descriptive analysis will be hold using the following tools:

5. For every numeric variable, a histogram and a boxplot are shown.
6. For the representation of the most technical variables, namely, ENS, TLP, RT and ETI a histogram of their logarithm are also performed.
7. When possible, regarding to the nature of the variable and its characteristics, a time series plot is shown and also, a plot showing the different values taken by the variable for each year of study.
8. For every categorical variable, a pie of frequencies and a barplot are shown.
9. For every variable of study, statistical indicators are studied as the mean, the median, the typical deviation or the frequency of appearance.

There are 48 original variables of study or columns. These variables includes the name of the country, the year and the month of occurrence and the station where it took place, used for identification. The variables in the dataset are:

- Year
- Month
- ID
- Country
- Station
- Reason
- ENS
- GDPperCap
- GINIindex
- FuelExp
- TotPop
- PopRur
- PopUrb
- EnergyInten
- CorrupRk
- DemocRk
- Nsub
- TLP
- RT
- ETI
- CO2Emis
- AccElecPop
- AccElecRur
- ElecProdHyd
- Climate
- Island
- ElecDistr
- ElecGen
- ElecTrans
- Eleccome
- RegPrice
- ElecProdNuc
- ElecProdOilGasCoal
- ElecProdRen
- EnerImp
- ElecPowCons
- TransDistrLoss
- Intercon
- Nuclear
- OCDE
- RatParis
- Government
- EU
- TempAverage
- PrepAverage
- Boperyear

Blackouts Characteristics Variables:

The number of blackouts registered for each year, which is equivalent to the frequency of appearance of every year in the dataset is shown.

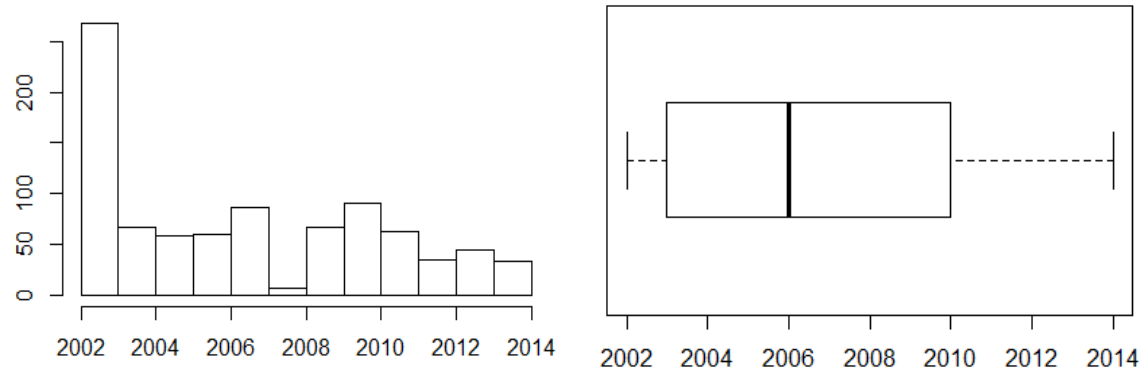


Figure 90

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2002    2003    2006    2007    2010    2014
## [1] "sd:  3.78173168035374"
## [1] "vc:  0.00188453260999493"
## [1] "Extended Summary Statistics"
```

- Country:

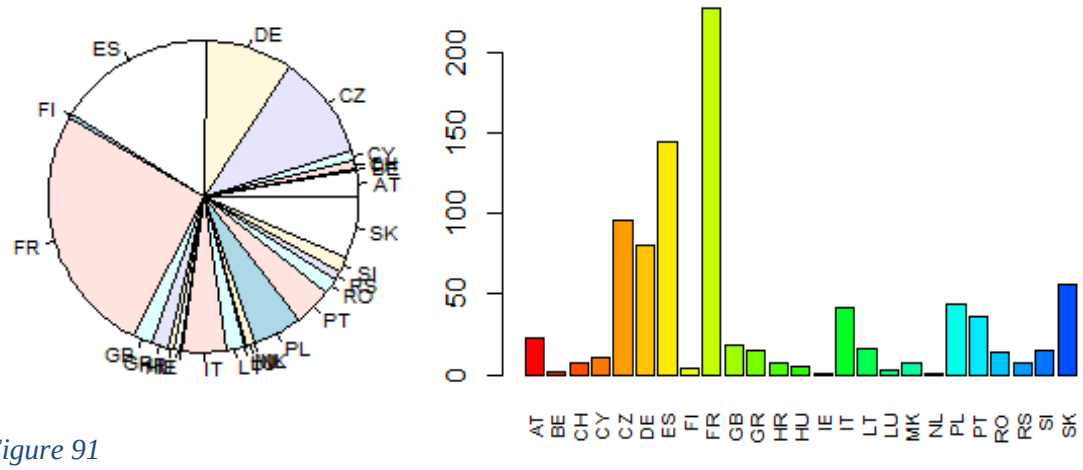


Figure 91

```
## [1] "Number of modalities:  25"
## [1] "Frequency table"
##  AT  BE  CH  CY  CZ  DE  ES  FI  FR  GB  GR  HR  HU  IE  IT  LT  LU  MK
##  23   2   7  10  96  80 144   4 227  18  15   7   5   1  41  16   3   7
##  NL  PL  PT  RO  RS  SI  SK
##   1  44  36  14   7  15  56
## [1] "Relative frequency table (proportions)"
##          AT          BE          CH          CY          CZ          DE
## 0.026166098 0.002275313 0.007963595 0.011376564 0.109215017 0.091012514
##          ES          FI          FR          GB          GR          HR
## 0.163822526 0.004550626 0.258248009 0.020477816 0.017064846 0.007963595
##          HU          IE          IT          LT          LU          MK
## 0.005688282 0.001137656 0.046643914 0.018202503 0.003412969 0.007963595
##          NL          PL          PT          RO          RS          SI
## 0.001137656 0.050056883 0.040955631 0.015927190 0.007963595 0.017064846
##          SK
## 0.063708760
## [1] "Frequency table sorted"
##
##  FR  ES  CZ  DE  SK  PL  IT  PT  AT  GB  LT  GR  SI  RO  CY  CH  HR  MK
## 227 144  96  80  56  44  41  36  23  18  16  15  15  14  10   7   7   7
##  RS  HU  FI  LU  BE  IE  NL
##   7   5   4   3   2   1   1
## [1] "Relative frequency table (proportions) sorted"
##          FR          ES          CZ          DE          SK          PL
## 0.258248009 0.163822526 0.109215017 0.091012514 0.063708760 0.050056883
##          IT          PT          AT          GB          LT          GR
## 0.046643914 0.040955631 0.026166098 0.020477816 0.018202503 0.017064846
##          SI          RO          CY          CH          HR          MK
## 0.017064846 0.015927190 0.011376564 0.007963595 0.007963595 0.007963595
##          RS          HU          FI          LU          BE          IE
## 0.007963595 0.005688282 0.004550626 0.003412969 0.002275313 0.001137656
##          NL
## 0.001137656
```

- Month:

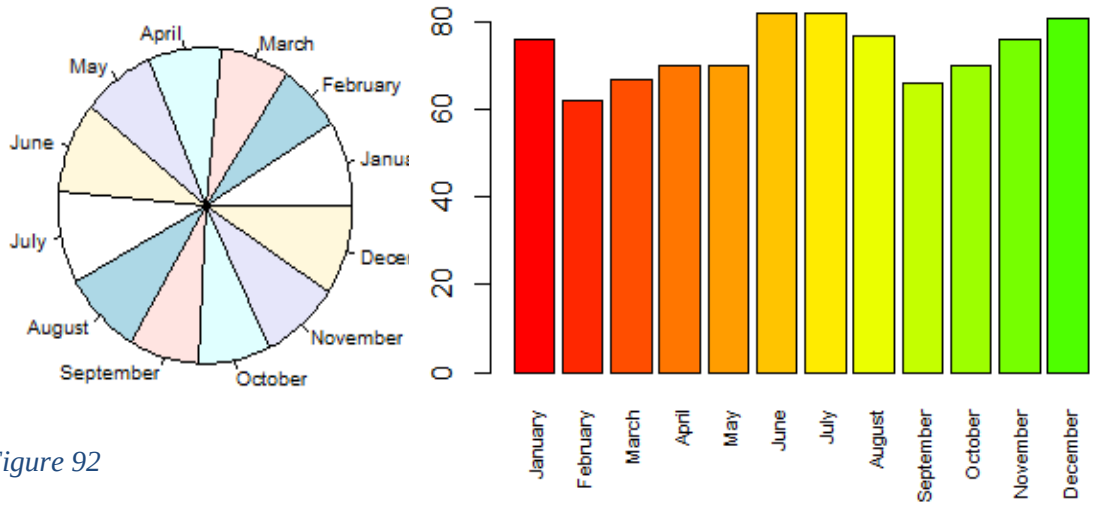


Figure 92

```
## [1] "Number of modalities:  12"

## [1] "Frequency table"
##   January  February     March     April       May      June      July 
##        76        62        67        70        70        82        82 
##    August September   October  November  December 
##        77        66        70        76        81 

## [1] "Relative frequency table (proportions)"
##    January   February      March      April        May       June 
## 0.08646189 0.07053470 0.07622298 0.07963595 0.07963595 0.09328783 
##       July     August  September    October   November   December 
## 0.09328783 0.08759954 0.07508532 0.07963595 0.08646189 0.09215017 

## [1] "Frequency table sorted"
##      June      July  December    August   January  November     April 
##        82        82        81        77        76        76        70 
##       May   October     March September  February 
##        70        70        67        66        62 

## [1] "Relative frequency table (proportions) sorted"
##       June       July   December     August    January   November 
## 0.09328783 0.09328783 0.09215017 0.08759954 0.08646189 0.08646189 
##      April        May    October      March  September   February 
## 0.07963595 0.07963595 0.07963595 0.07622298 0.07508532 0.07053470
```

- Reason:

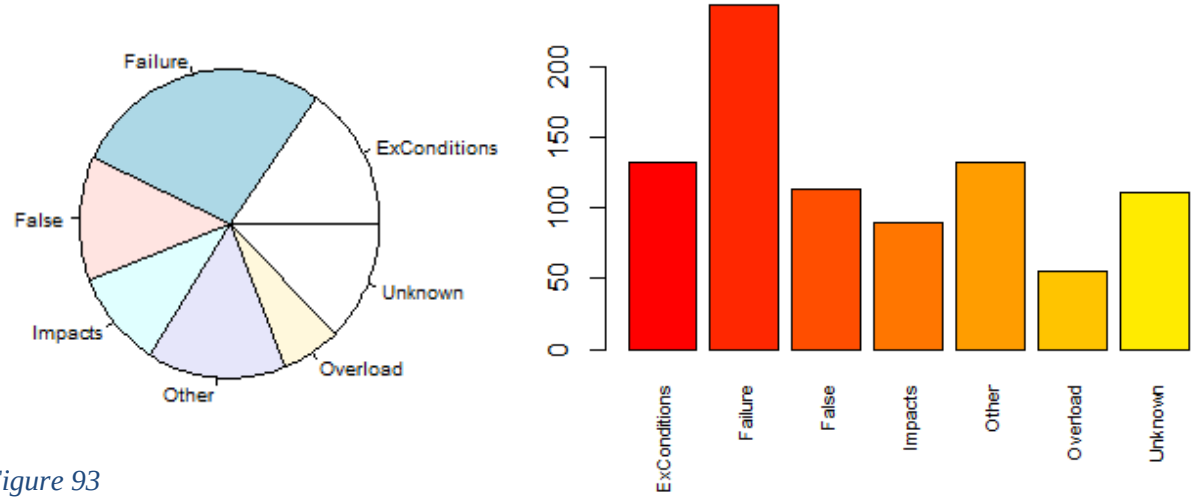
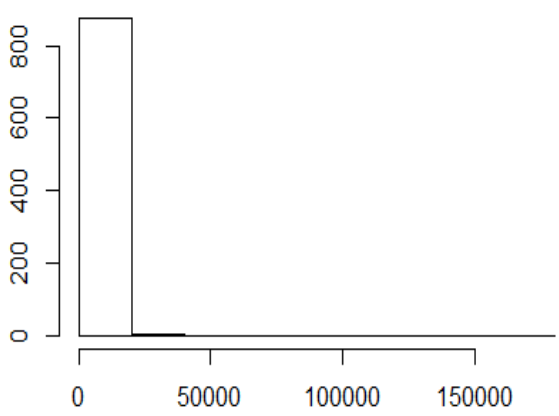
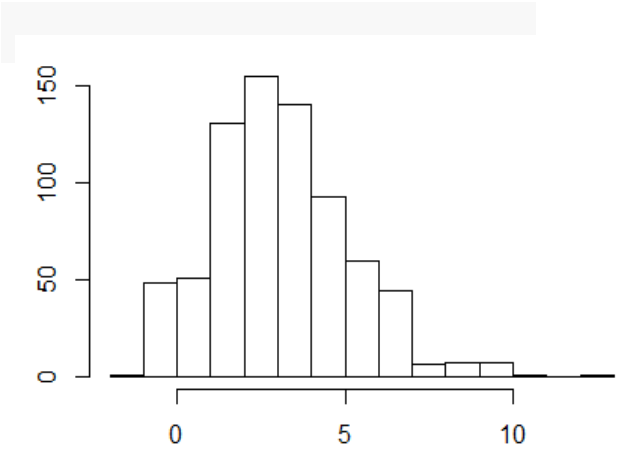


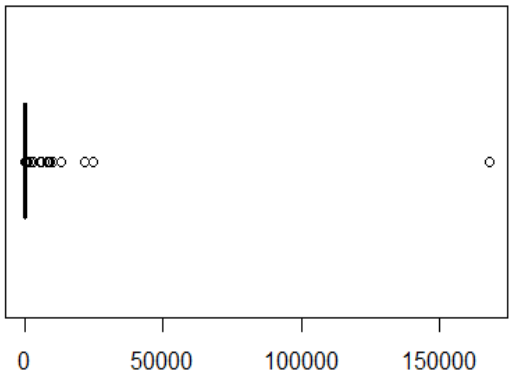
Figure 93

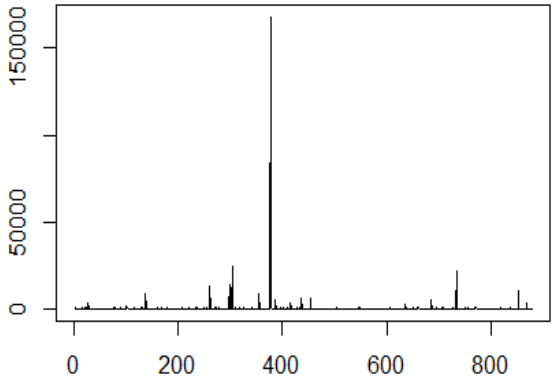
```
## [1] "Number of modalities:  7"
## [1] "Frequency table"
## ExConditions      Failure        False      Impacts        Other
##          133          244          114           90          132
##     Overload      Unknown
##           55          111
## [1] "Relative frequency table (proportions)"
## ExConditions      Failure        False      Impacts        Other
##    0.1513083    0.2775882    0.1296928    0.1023891    0.1501706
##     Overload      Unknown
##    0.0625711    0.1262799
## [1] "Frequency table sorted"
##      Failure ExConditions        Other        False      Unknown
##          244          133          132          114          111
##      Impacts     Overload
##           90           55
## [1] "Relative frequency table (proportions) sorted"
##      Failure ExConditions        Other        False      Unknown
##    0.2775882    0.1513083    0.1501706    0.1296928    0.1262799
##      Impacts     Overload
##    0.1023891    0.0625711
```

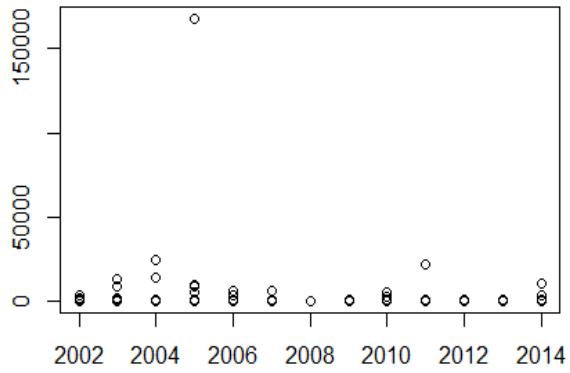
- Energy Not Supplied during the Failure:







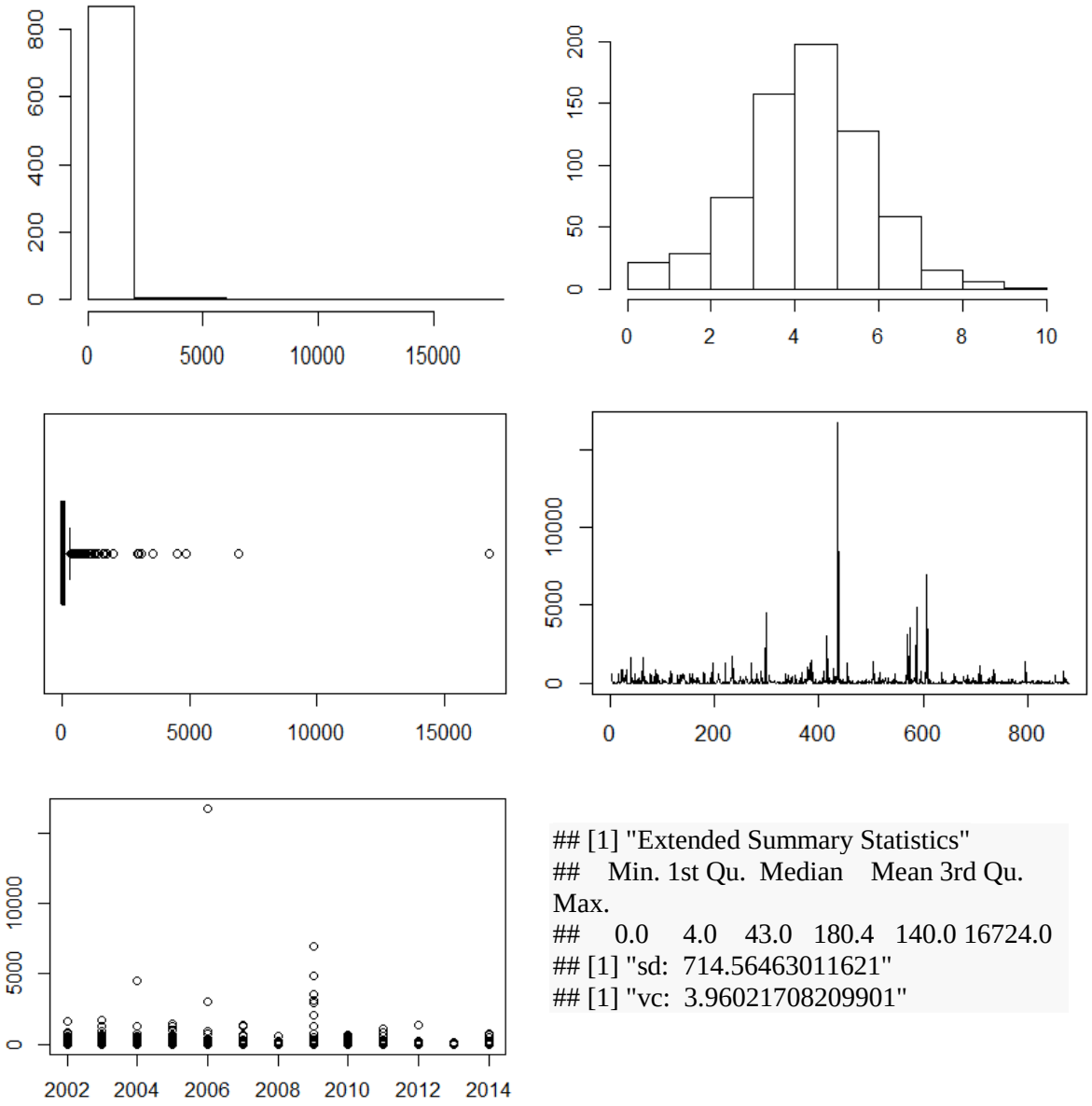





```
## [1] "Extended Summary Statistics"
##     Min.  1st Qu.   Median     Mean  3rd
Qu.     Max.
##      0.0      2.0     12.0    434.8     54.0 168
000.0
## [1] "sd:  5850.7901230051"
## [1] "vc:  13.4560290565908"
```

Figure 94

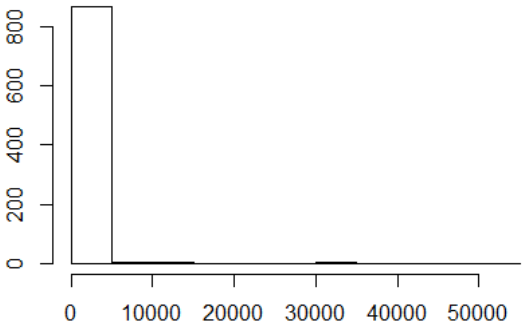
- Total Loss of Power during the Failure:

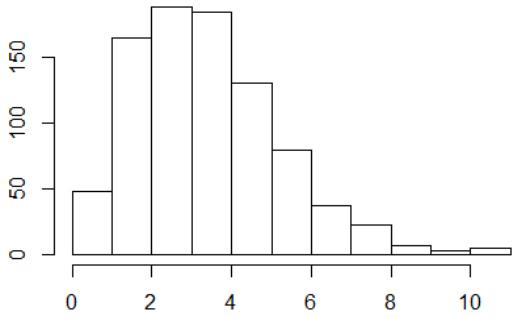


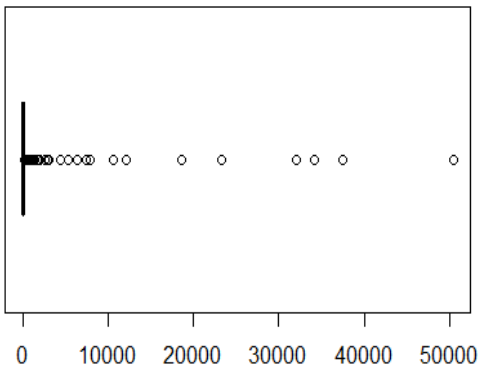

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##     0.0     4.0    43.0   180.4   140.0 16724.0
## [1] "sd:  714.56463011621"
## [1] "vc:  3.96021708209901"
```

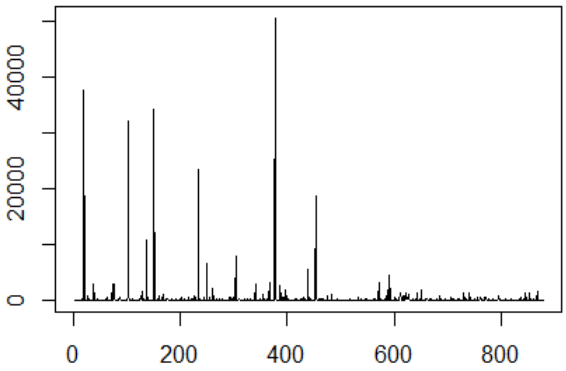
Figure 95

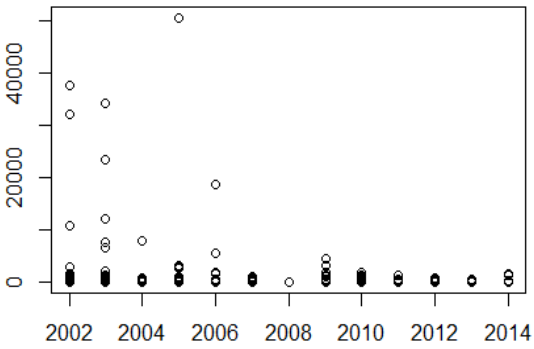
- Time Required to Restore the Service after a Failure:







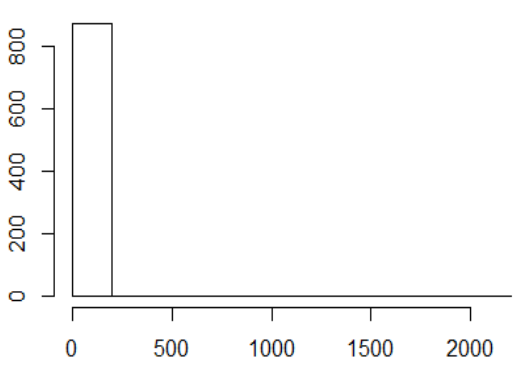


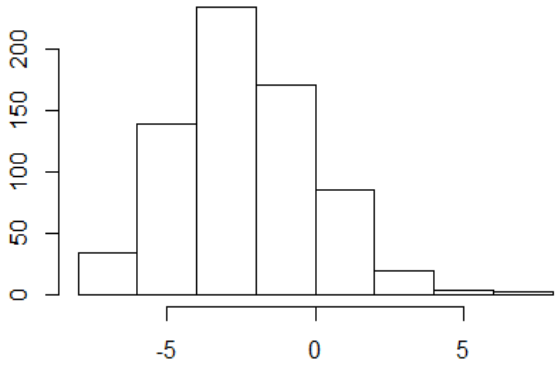


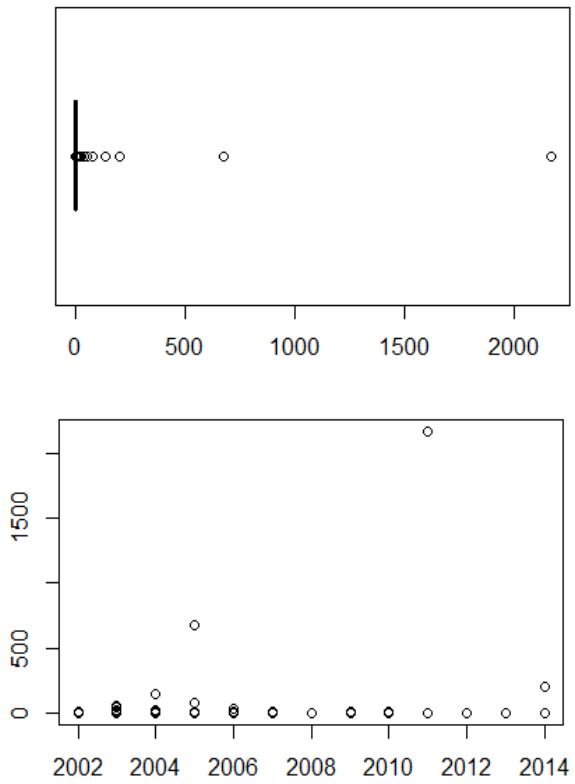

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##     0.0     7.0    24.0   412.3    82.0 50432.
0
## [1] "sd:  2919.26049771817"
## [1] "vc:  7.08118424246363"
```

Figure 96

- Equivalent Time of the Blackout Taking into Account the Amount of Service during 12 Months:








```
## [1] "Extended Summary Statistics"
##      Min.  1st Qu.   Median     Mean  3rd Qu.
Max.     NA's
##    0.000    0.002    0.039    4.521    0.230 217
0.716        4
## [1] "sd:  77.3862294502699"
## [1] "vc:  17.1174462401342"
```

Figure 97

Economic variables:

- GDP per capita:

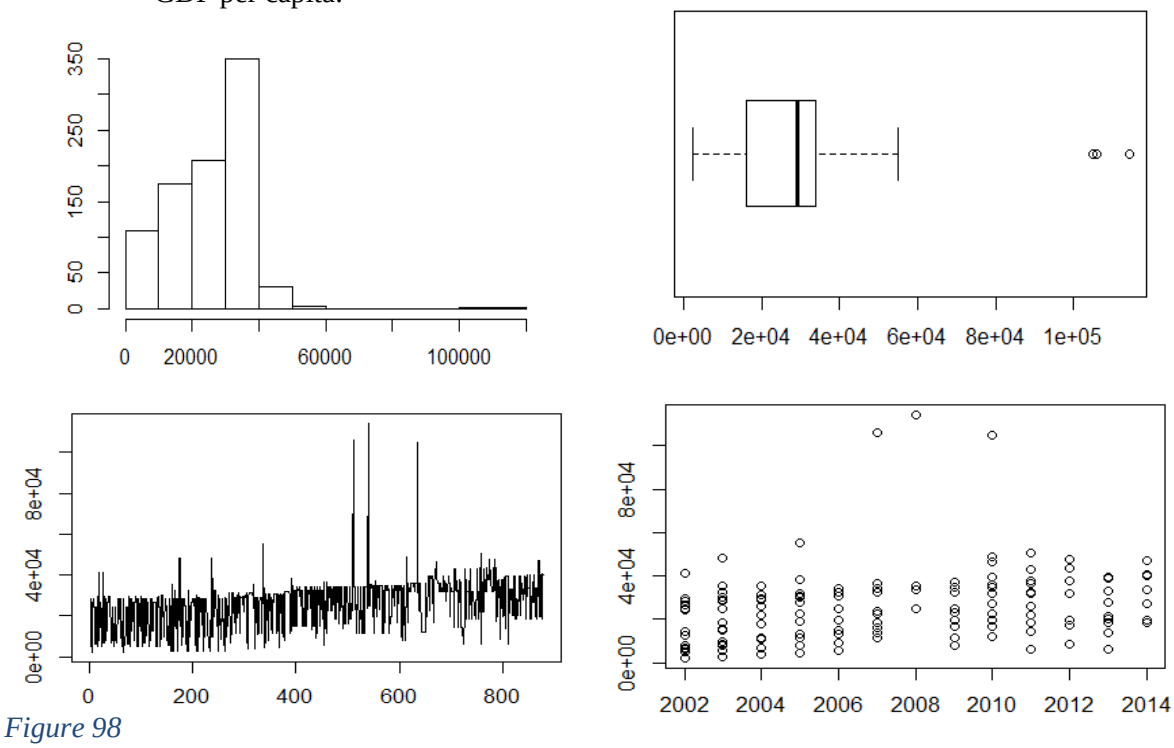


Figure 98

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2150   16058   29080   25849   33822  114294
## [1] "sd:  11783.1627445441"
## [1] "vc:  0.455844763574785
```

- GINI Index:

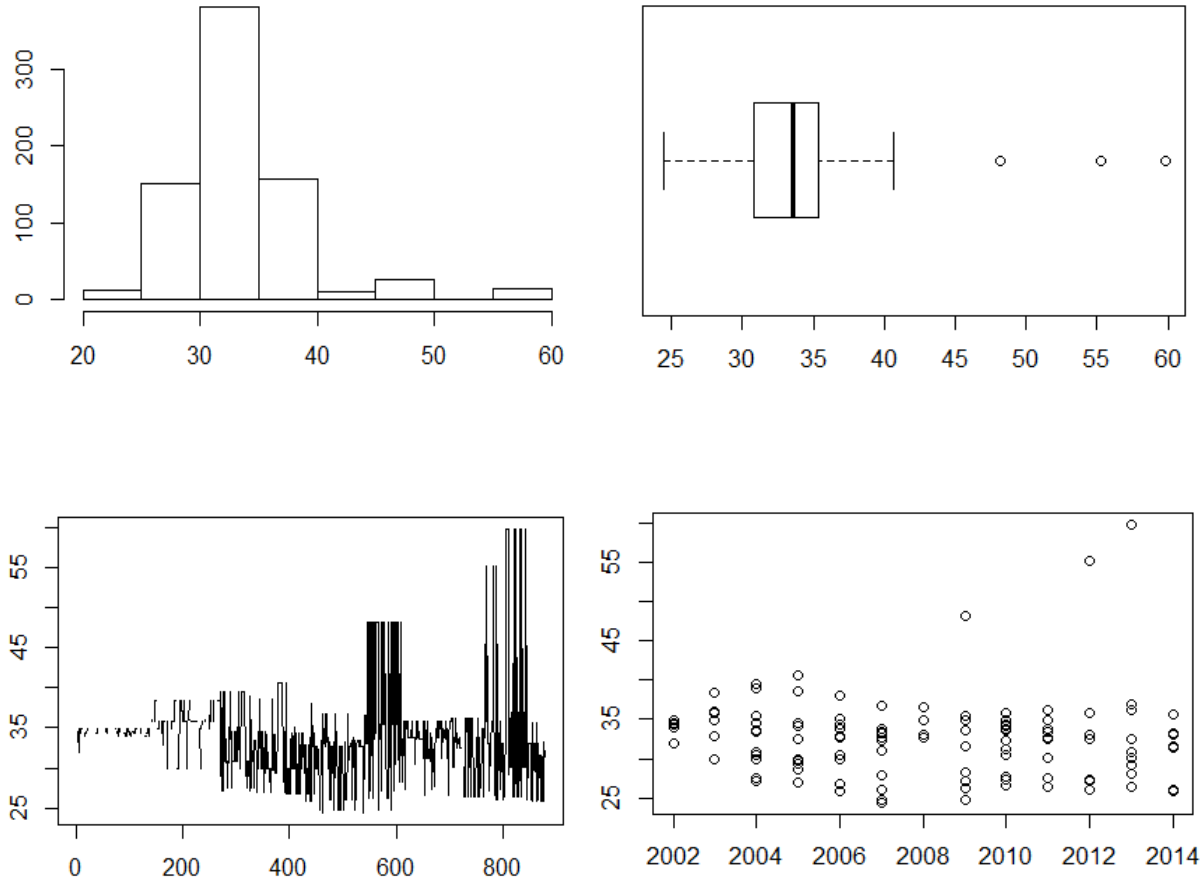


Figure 99

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   24.40   30.80   33.60   33.67   35.40   59.80     128
## [1] "sd:  5.64095229673375"
## [1] "vc:  0.167531349180527”
```

- Total Population:

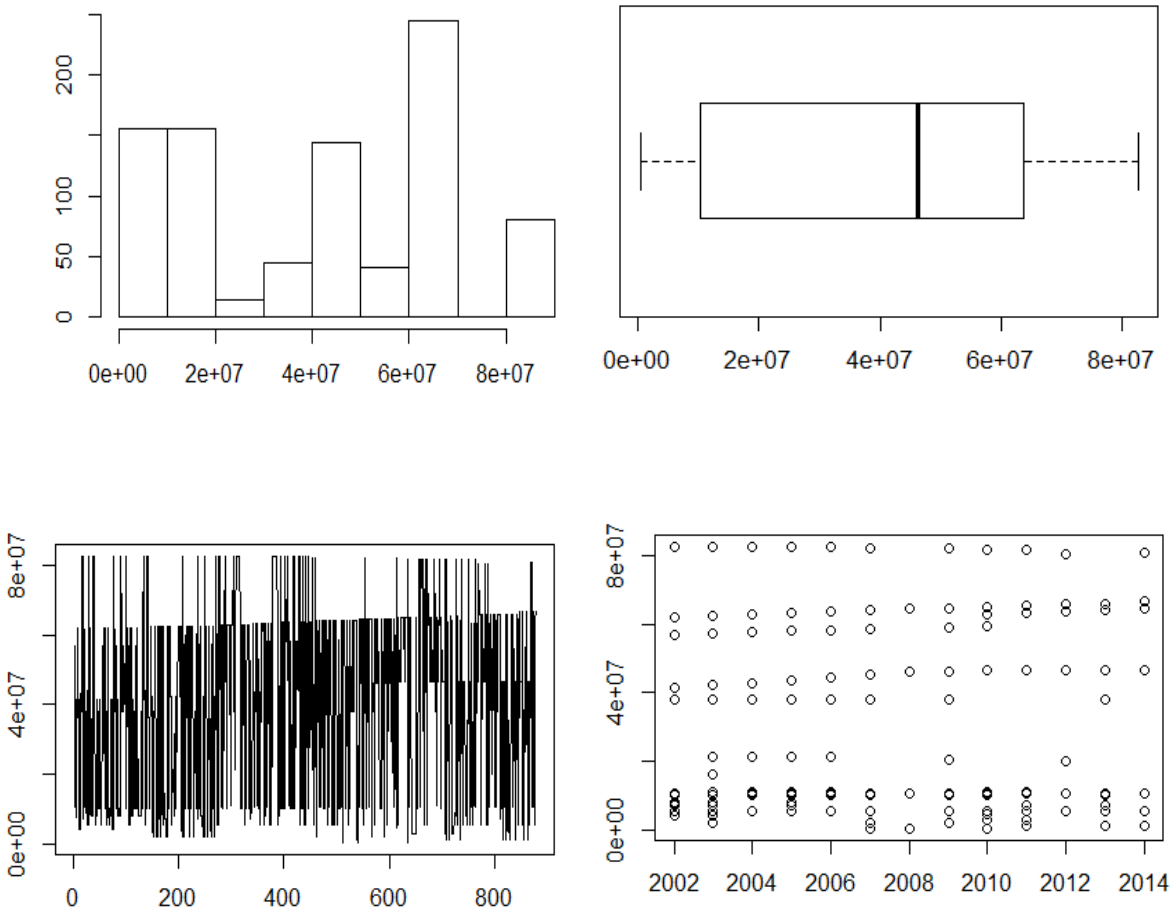
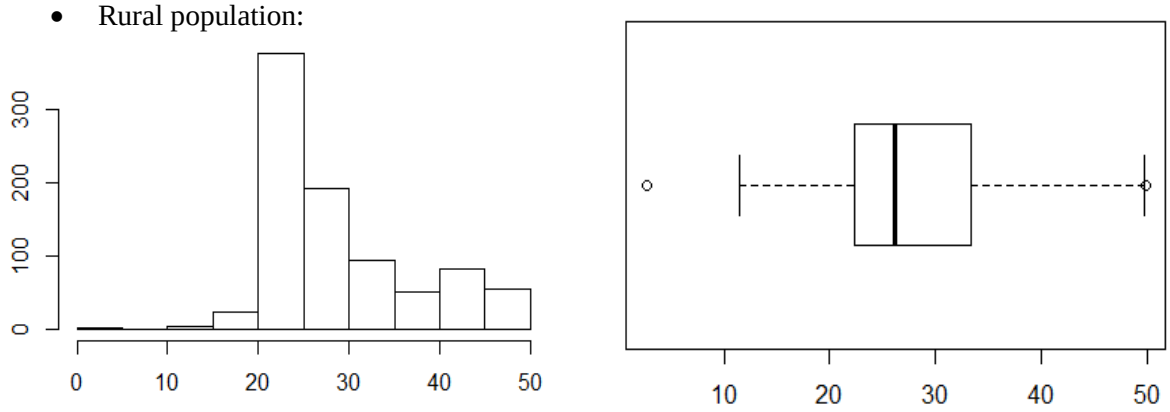
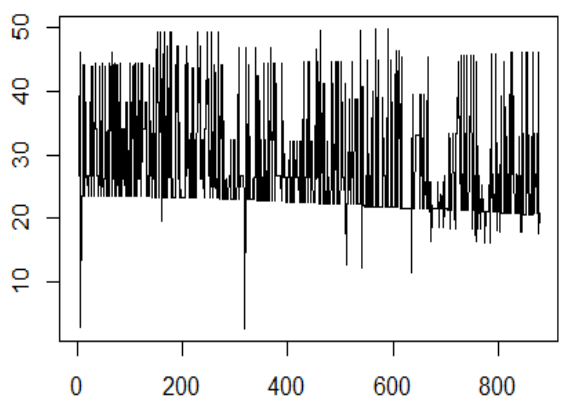


Figure 100

```
## [1] "Extended Summary Statistics"
##    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  479993 10298828 46362946 40285600 63621376 82534176
## [1] "sd:  26868123.1621744"
## [1] "vc:  0.666941124678497"
```

- Rural population:





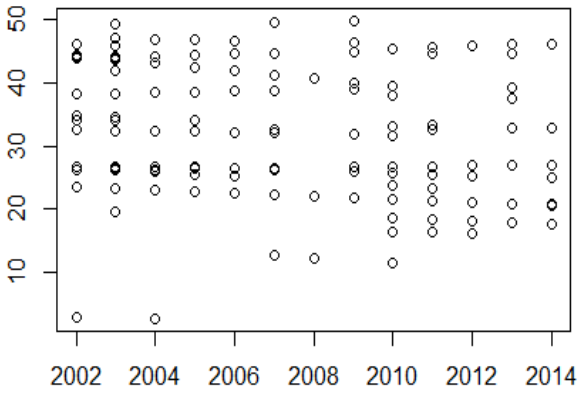
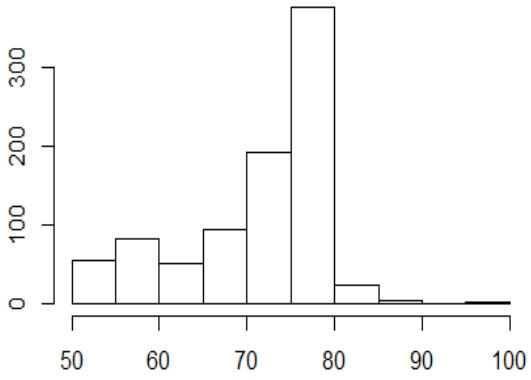
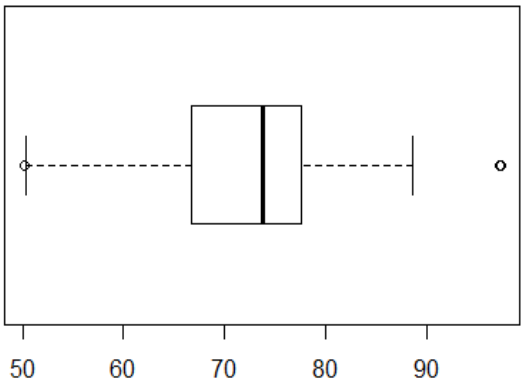


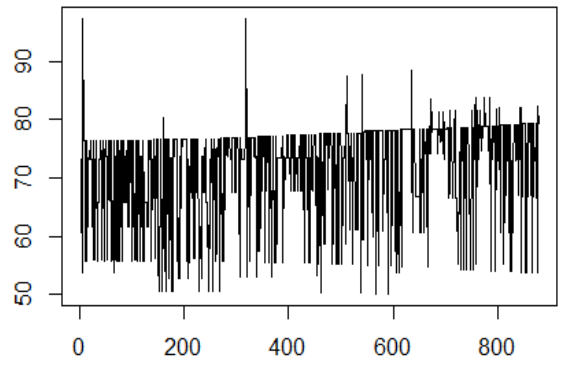
Figure 101

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##   2.655  22.379  26.191  28.616  33.328  49.867
## [1] "sd:  8.62487087933485"
## [1] "vc:  0.301395910799084"
```

- Urban Population:







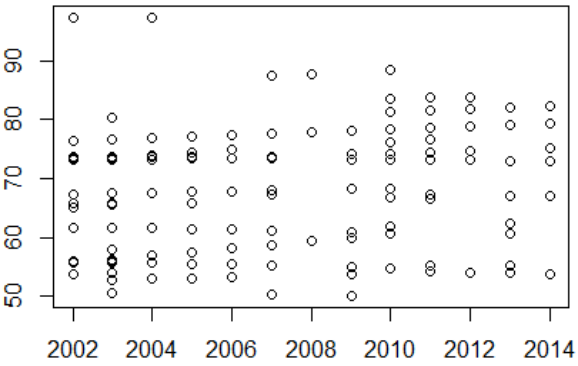


Figure 102

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##   50.13   66.67   73.81   71.38   77.62   97.34
## [1] "sd:  8.62487087933485"
## [1] "vc:  0.120824290995535"
```

- Corruption Ranking:

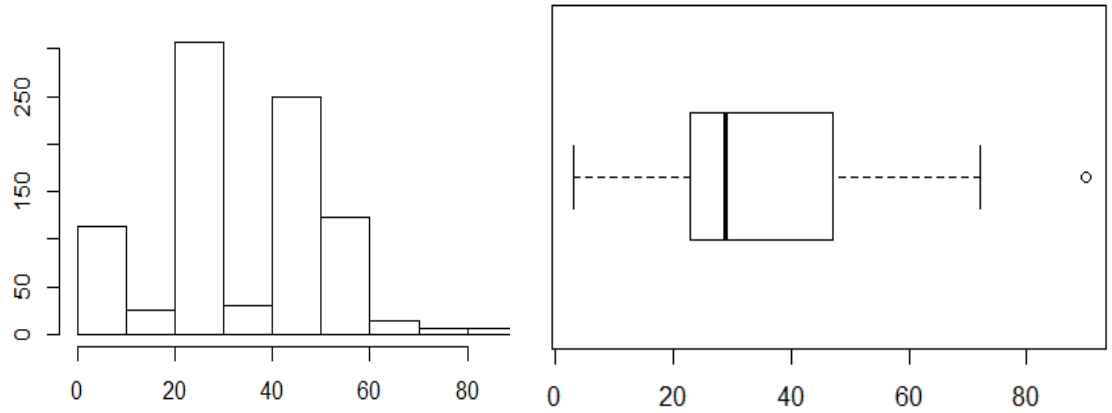


Figure 103

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00   23.00   29.00   34.27   47.00   90.00
## [1] "sd:  16.691917996804"
## [1] "vc:  0.487027681045964"
## [1] "variable  DemocRk : DemocRk"
```

- Democracy Ranking:

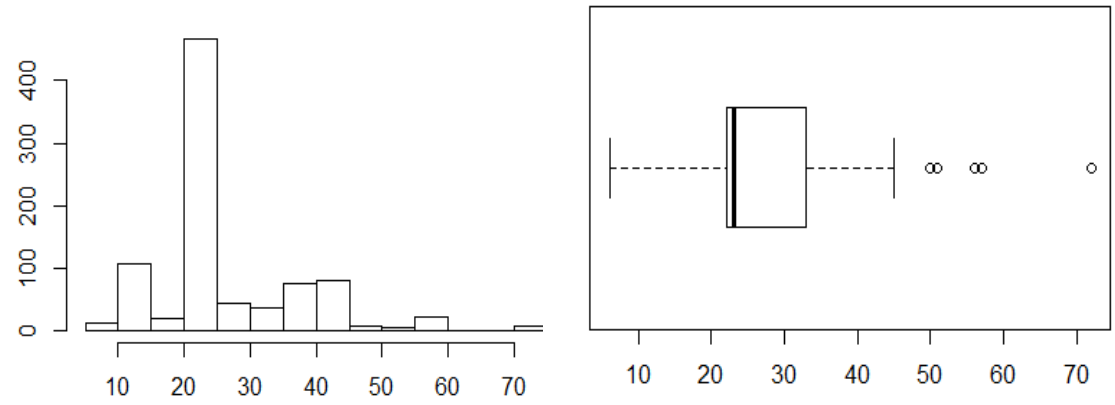


Figure 104

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.00   22.00   23.00   27.05   33.00   72.00
## [1] "sd:  11.1973947303216"
## [1] "vc:  0.414003111296067"
```

- Belonging to the OCDE:

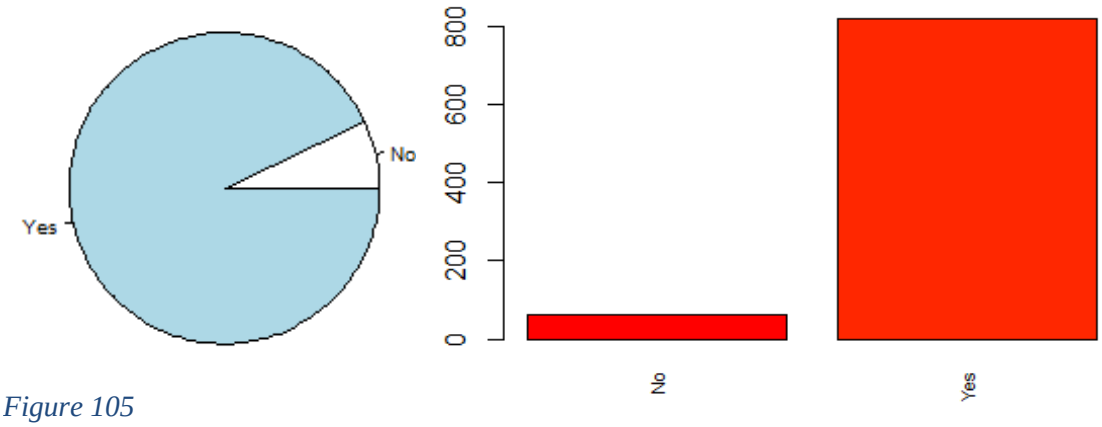


Figure 105

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
##  61 818
## [1] "Relative frequency table (proportions)"
##         No        Yes
## 0.06939704 0.93060296
## [1] "Frequency table sorted"
## Yes  No
## 818  61
## [1] "Relative frequency table (proportions) sorted"
##        Yes         No
## 0.93060296 0.06939704
```

- Type of Government:

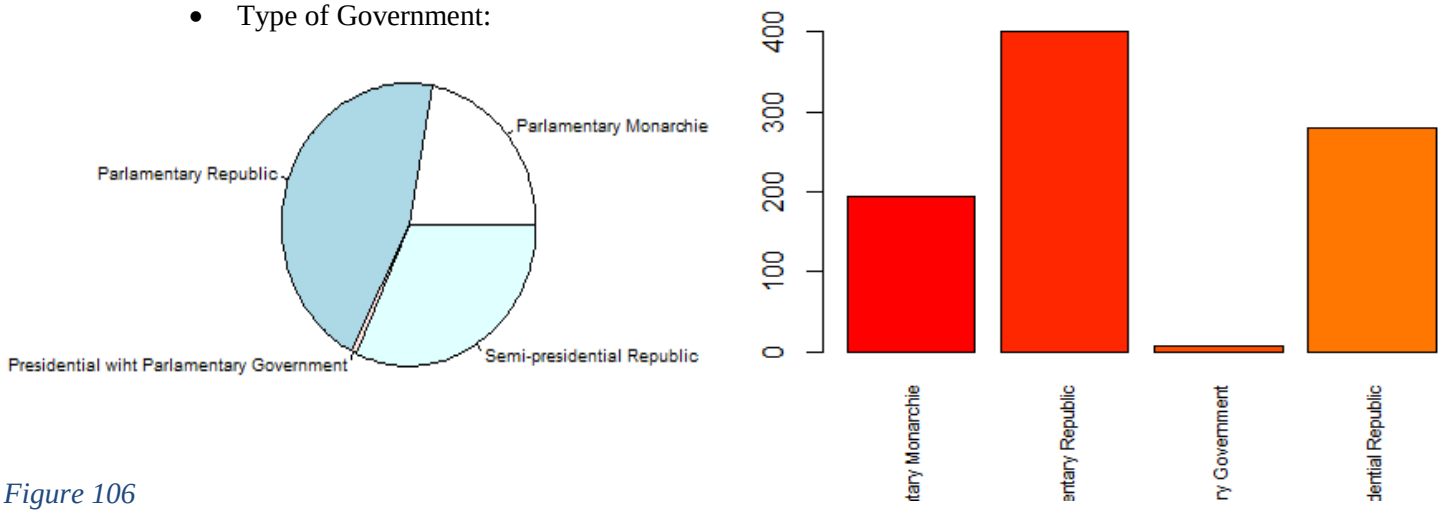


Figure 106

```
## [1] "Number of modalities:  4"
## [1] "Frequency table"
##                  Parlamentary Monarchie
##                                     193
##                   Parlamentary Republic
```

```
##                                400
## Presidential wiht Parlamentary Government
##                                  7
##               Semi-presidential Republic
##                                279
## [1] "Relative frequency table (proportions)"
##                    Parlamentary Monarchie
##                               0.219567691
##                    Parlamentary Republic
##                               0.455062571
## Presidential wiht Parlamentary Government
##                               0.007963595
##               Semi-presidential Republic
##                               0.317406143
## [1] "Frequency table sorted"
##                    Parlamentary Republic
##                                400
##               Semi-presidential Republic
##                                279
##                    Parlamentary Monarchie
##                                193
## Presidential wiht Parlamentary Government
##                                  7
## [1] "Relative frequency table (proportions) sorted"
##
##                    Parlamentary Republic
##                               0.455062571
##               Semi-presidential Republic
##                               0.317406143
##                    Parlamentary Monarchie
##                               0.219567691
## Presidential wiht Parlamentary Government
##                               0.007963595
```

- Belonging to the EU:

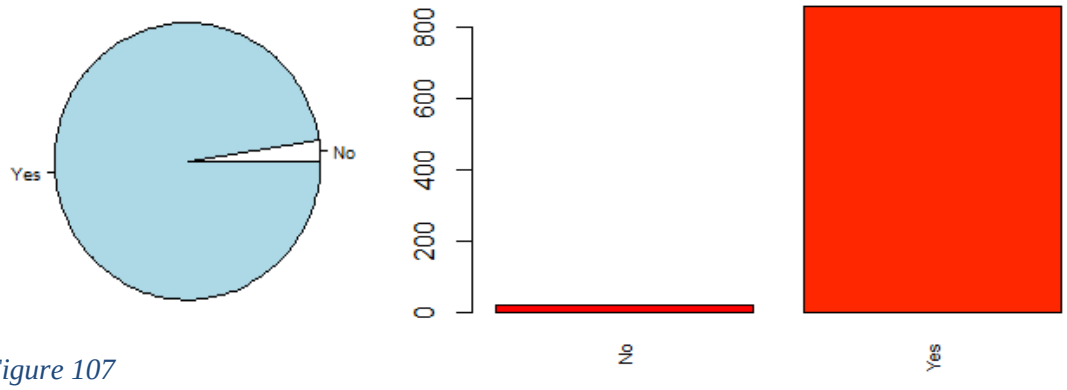


Figure 107

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
##  21 858
```

```
## [1] "Relative frequency table (proportions)"
##         No        Yes
## 0.02389078 0.97610922
## [1] "Frequency table sorted"
## Yes  No
## 858  21
## [1] "Relative frequency table (proportions) sorted"
##        Yes         No
## 0.97610922 0.02389078
```

Energy variables:

- Emissions of CO_2 ;

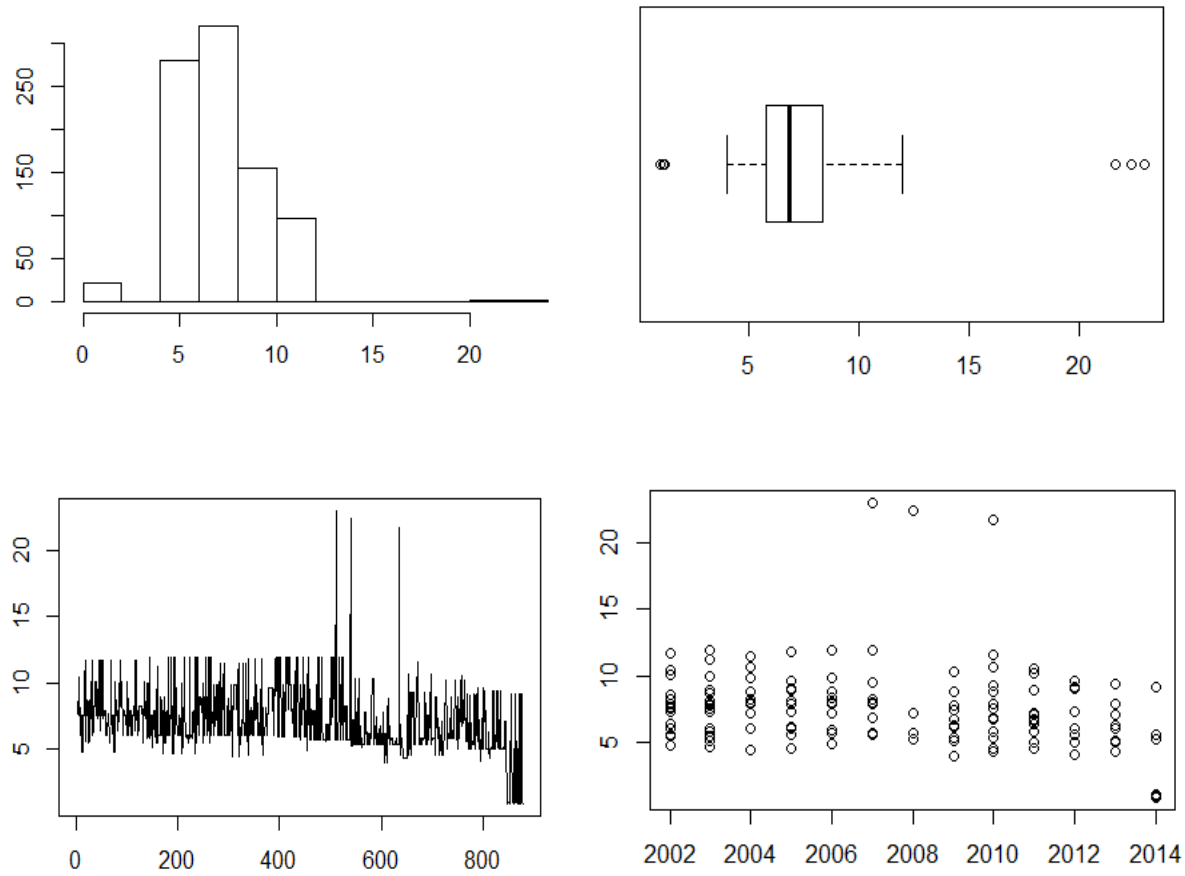


Figure 108

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.9197  5.7628  6.7997  7.2241  8.2991 22.9573       3
## [1] "sd:  2.3852114803841"
## [1] "vc:  0.330174891556975"
```

- Access to Electricity in terms of Total Population:

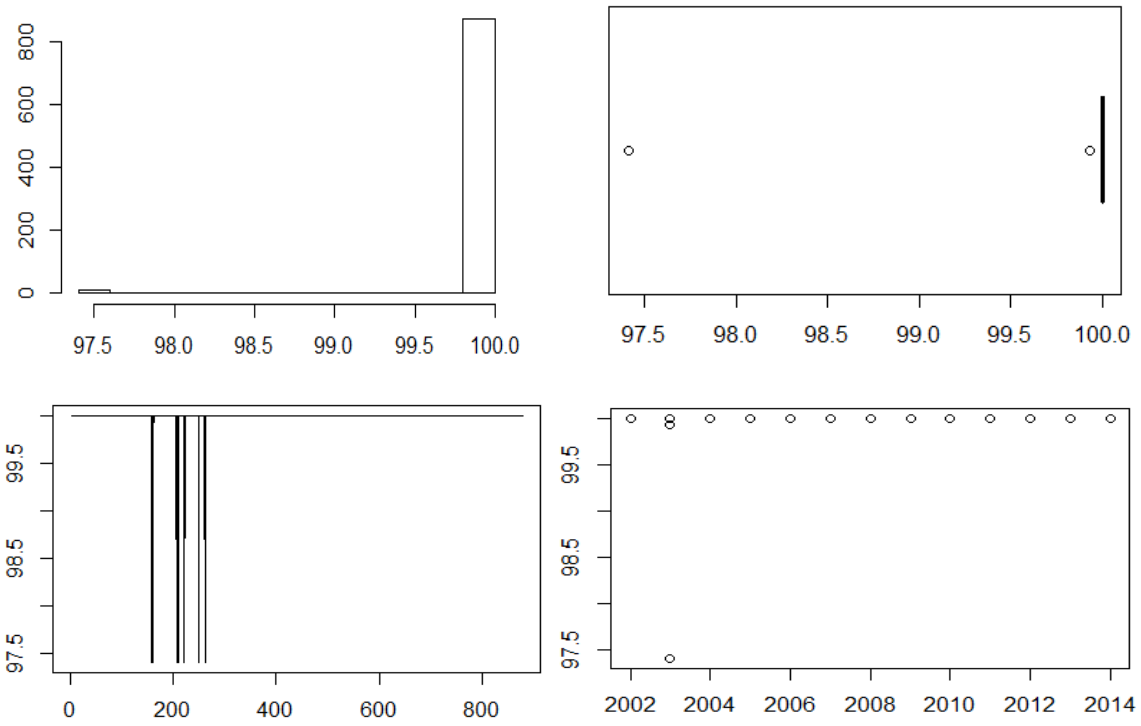


Figure 109

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   97.41  100.00  100.00   99.98  100.00  100.00
## [1] "sd:  0.230443977242402"
## [1] "vc:  0.00230491711526061"
```

- Access to Electricity in terms of Rural Population:

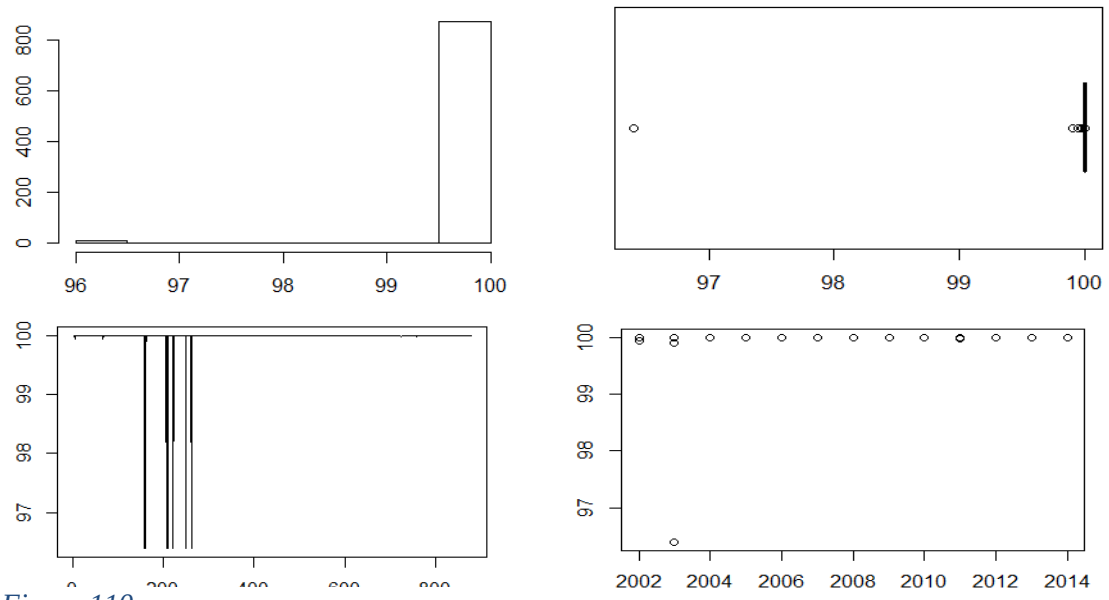


Figure 110

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   96.39  100.00  100.00   99.97  100.00  100.00
## [1] "sd:  0.320647243375942"
## [1] "vc:  0.003207403153503"
```

- Electricity Production from Hydrological Sources:

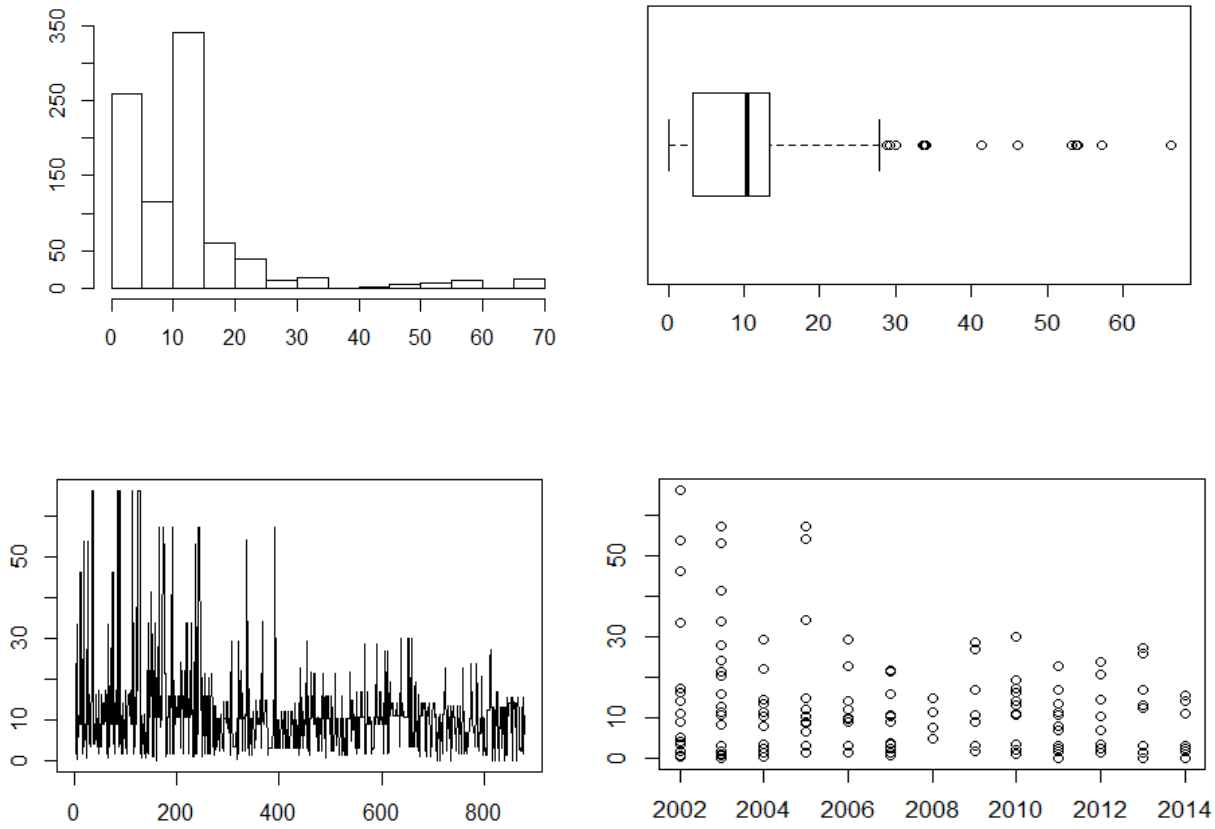


Figure 111

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Ma
x.
##   0.000   3.288  10.472  11.724  13.408  66.265
## [1] "sd:  11.6336384610587"
## [1] "vc:  0.992288715082398”
```

- Electricity Produced from Nuclear Sources:

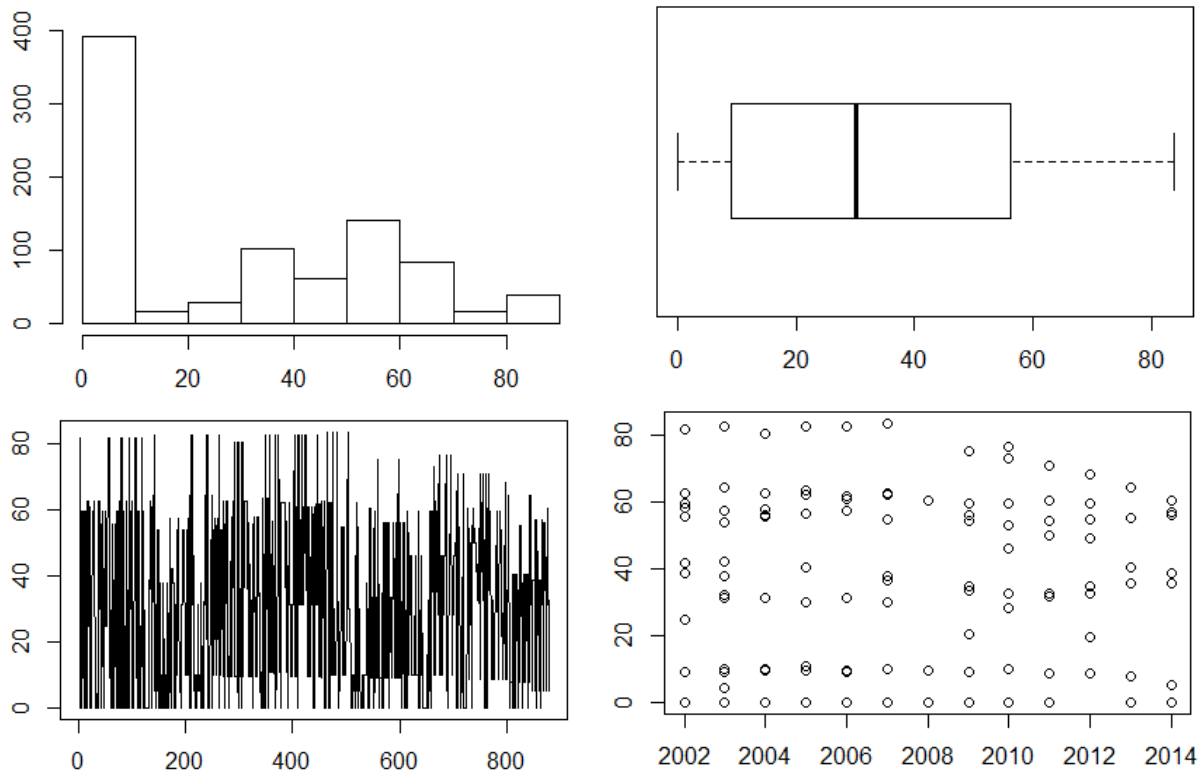
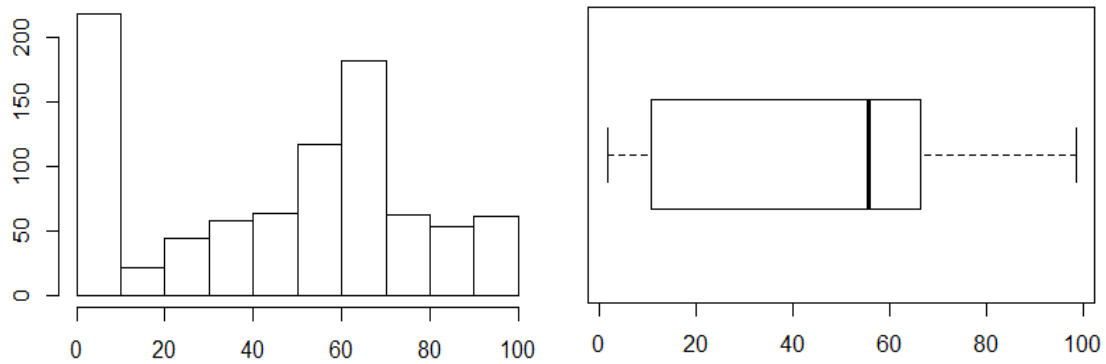
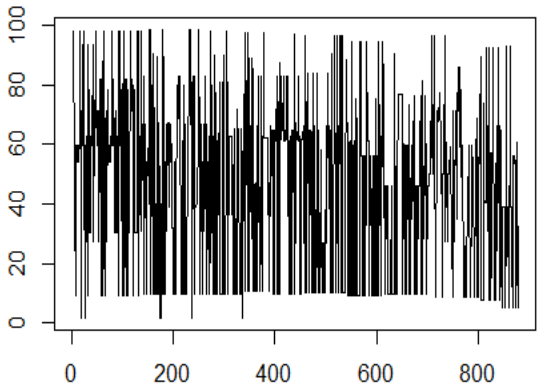


Figure 112

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    M
ax.
##   0.000   9.235  30.181  30.515  56.164  83.631
## [1] "sd:  26.0764403490368"
## [1] "vc:  0.854544418644029"
```

- Electricity Production from Oil, Gas and Coal Sources:





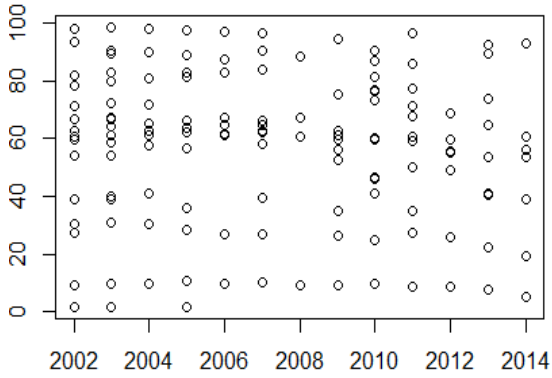
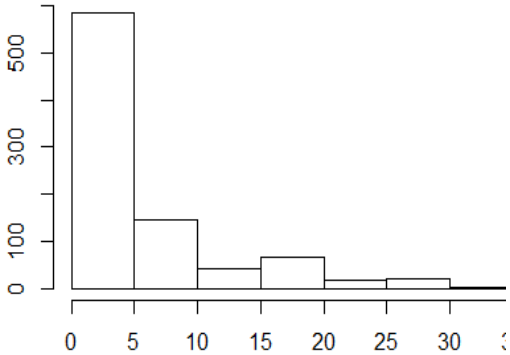
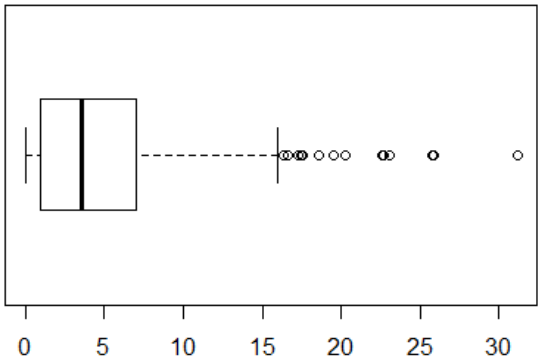


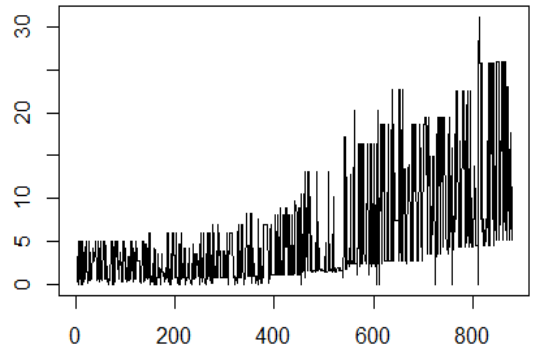
Figure 113

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    M
ax.
##   1.703  10.798  55.722  47.376  66.282  98.480
## [1] "sd:  28.1878261393327"
## [1] "vc:  0.594976553713286"
```

- Electricity Produced from Renewable Sources:







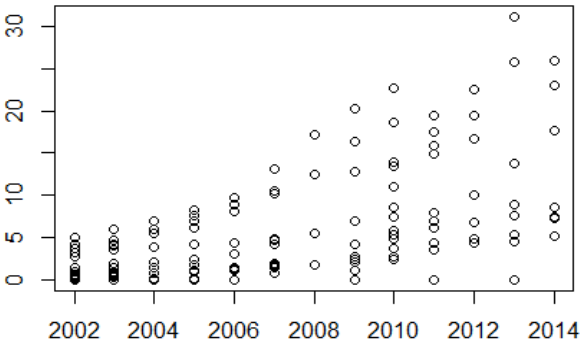


Figure 114

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.9395  3.5616  5.6266  6.9830 31.1493
## [1] "sd:  6.4654940341569"
## [1] "vc:  1.14910010838016"
```

- Energy Imports:

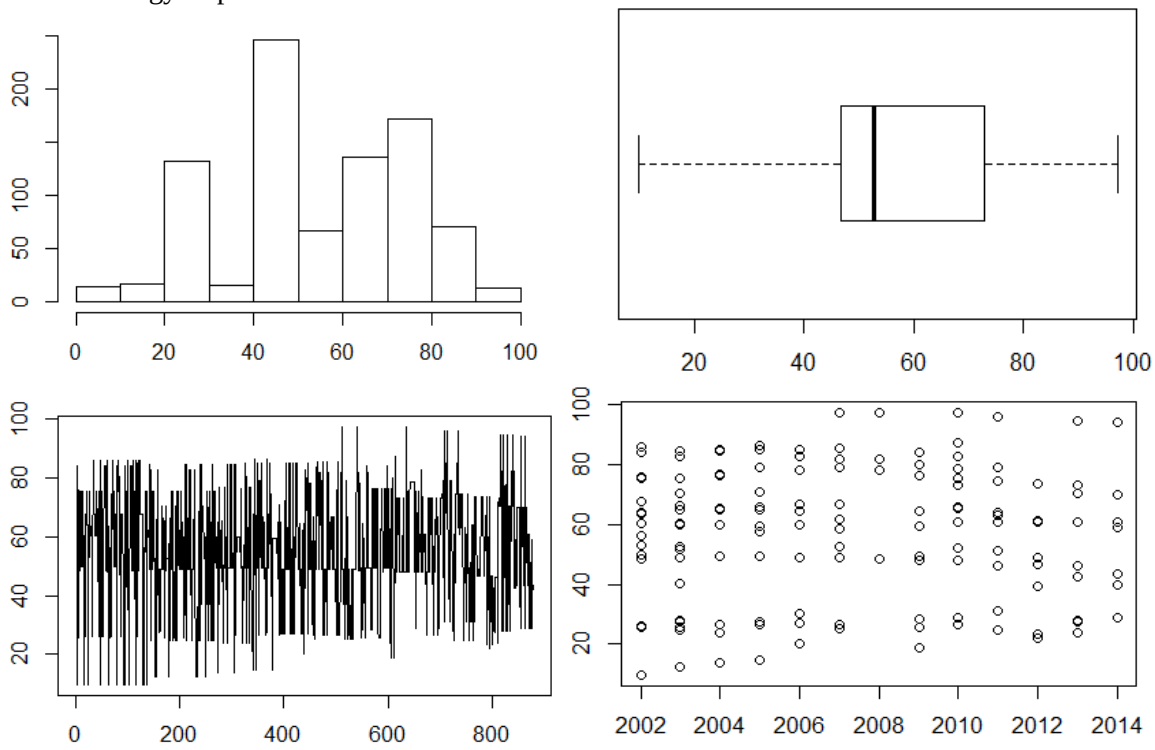
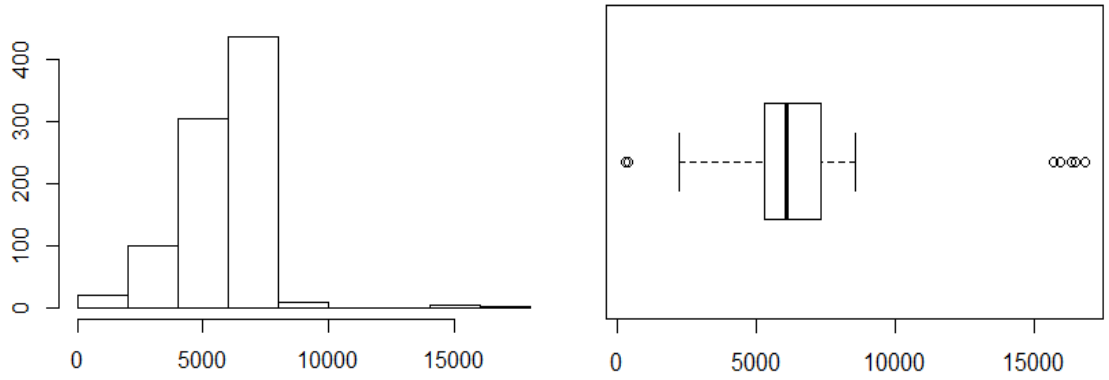


Figure 115

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   9.791  46.704  52.757  55.238  73.050  97.204
## [1] "sd:  19.9751637265829"
## [1] "vc:  0.36162001880523"
```

- Electric Power Consumed:




```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   263.6  5272.4  6074.8  5995.6  7340.7 16830.0
## [1] "sd:  1872.37694608688"
## [1] "vc:  0.312289397589361"
```

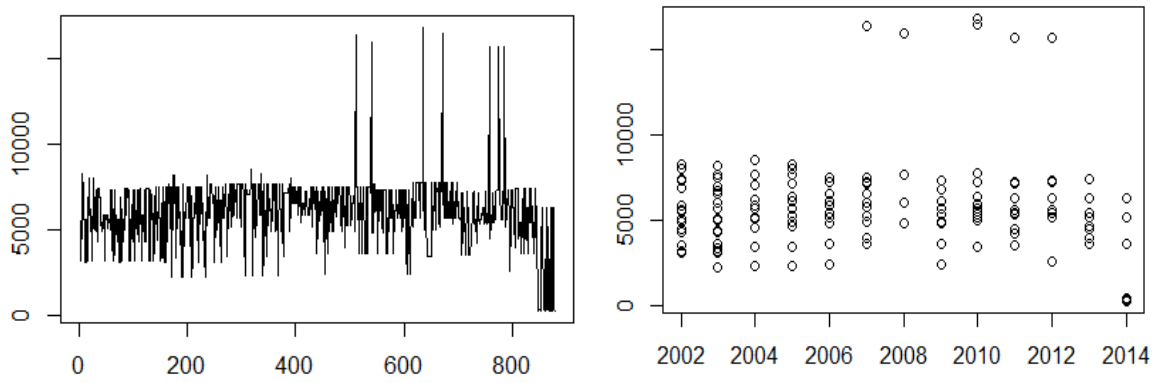
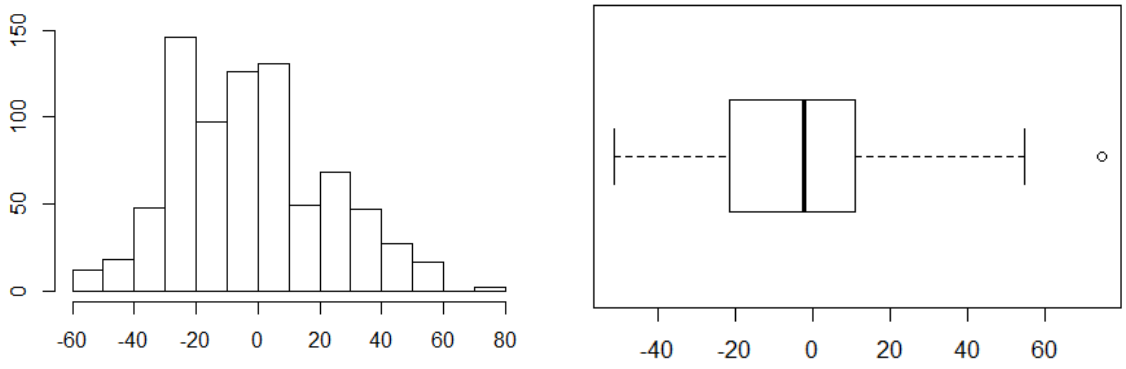


Figure 116

- Total Greenhouse Gas Emissions Change from 1990:



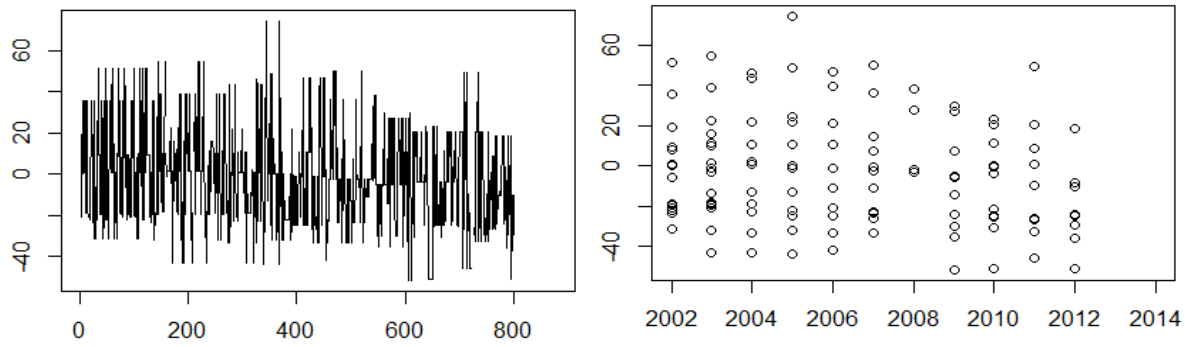


Figure 117

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.    NA's
## -51.614 -21.571  -2.477  -2.142  10.879  74.479      91
## [1] "sd:  24.1505459507512"
## [1] "vc:  -11.274130044175"
```

- Electric Power Transmission and Distribution Losses:

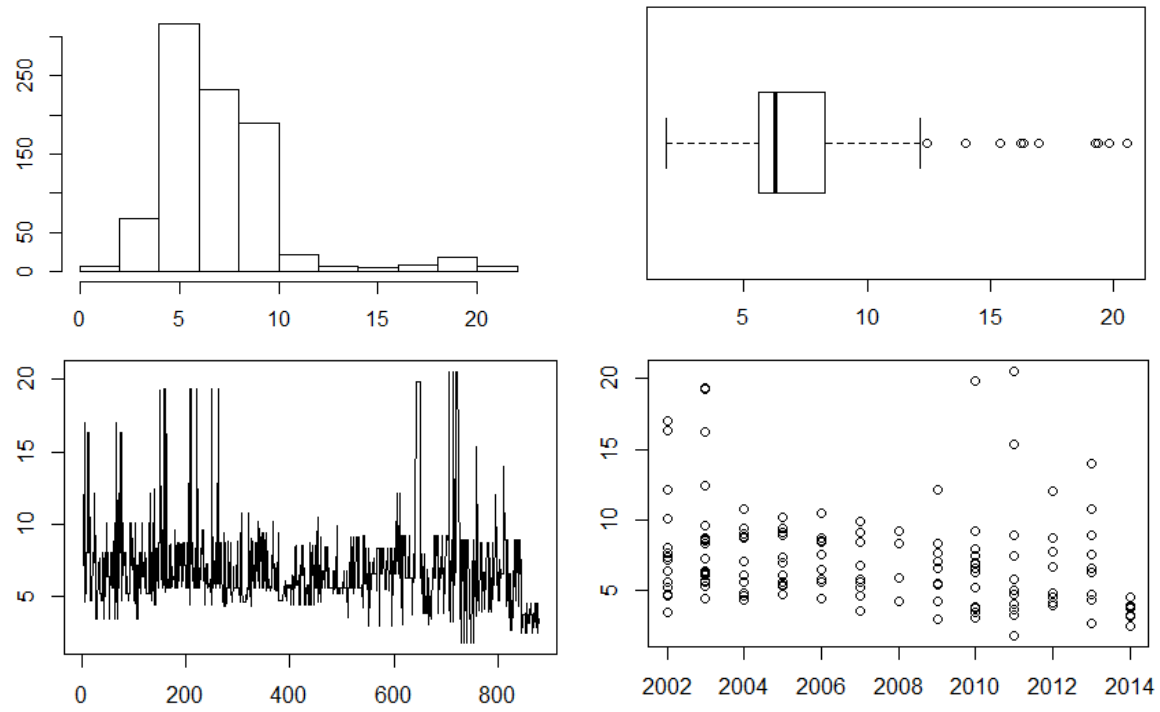


Figure 118

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.821   5.580   6.274   6.955   8.300  20.537
## [1] "sd:  3.07731197177107"
## [1] "vc:  0.442455118651542"
```

- Fuel Exports:

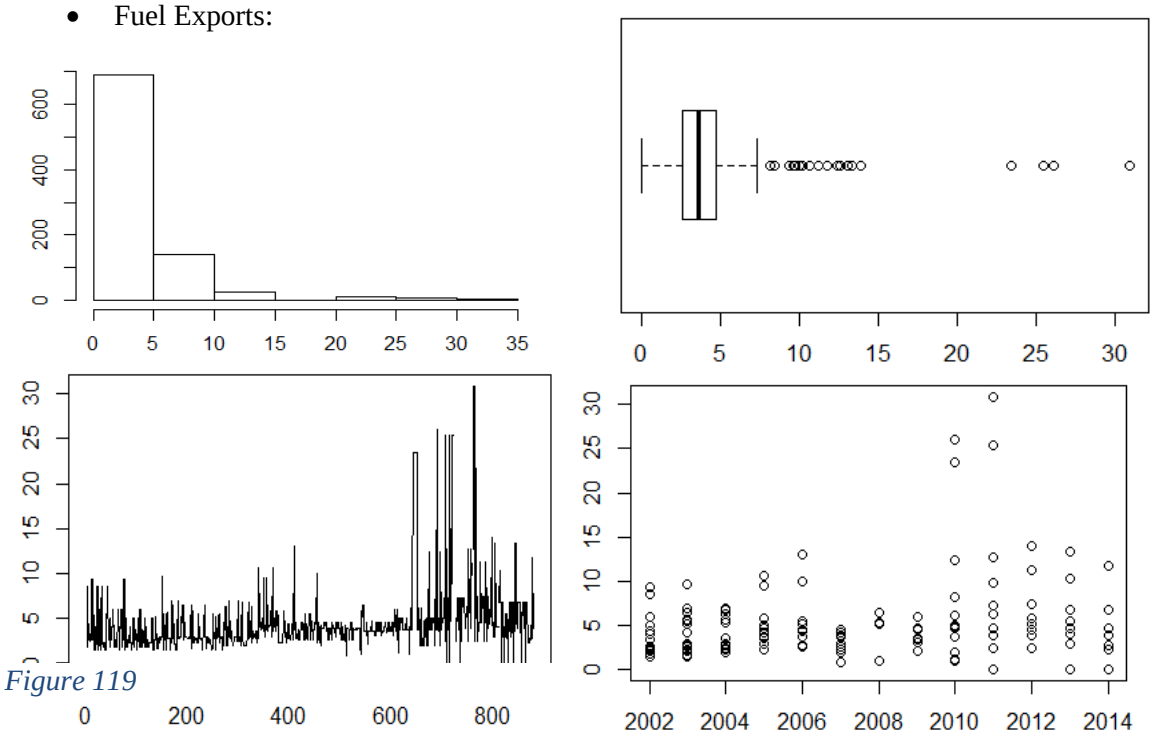


Figure 119

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   2.579   3.659   4.442   4.765  30.873       5
## [1] "sd:  3.77988117940105"
## [1] "vc:  0.850905996538367"
```

- Time required to get electricity:

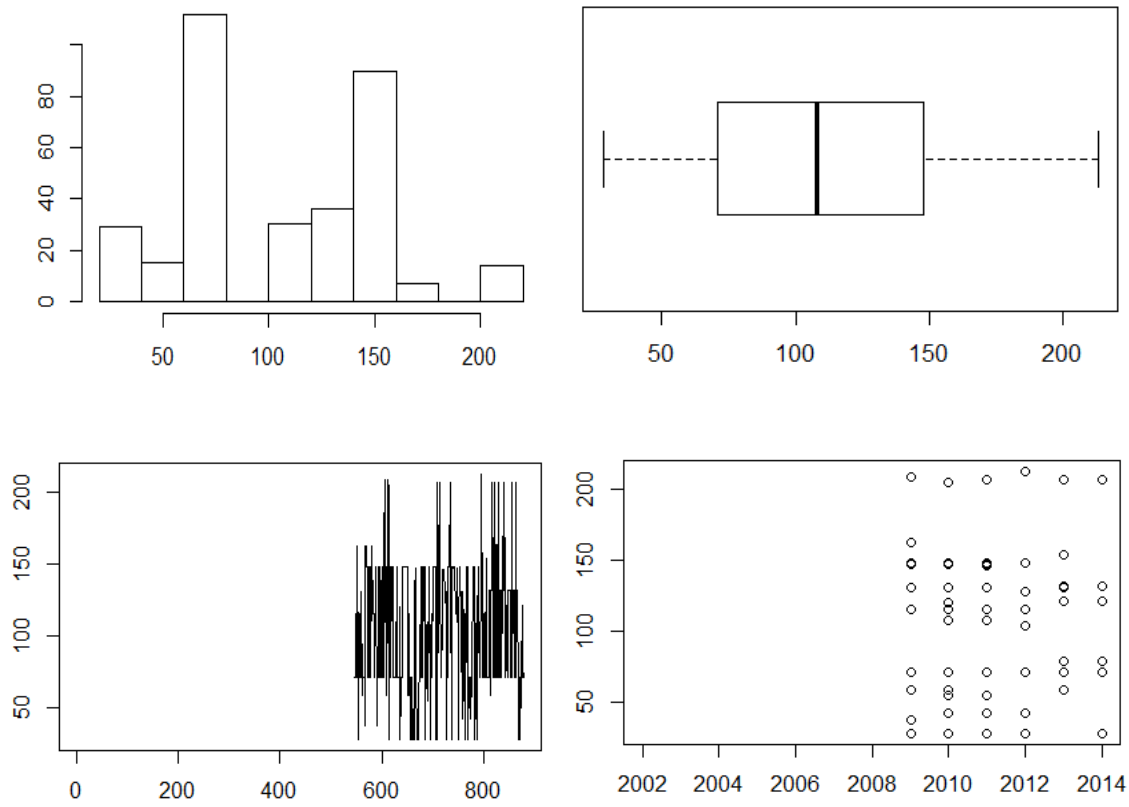


Figure 120

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    28.0    71.0   108.0   105.1   148.0   213.0     546
## [1] "sd:  45.8036475704421"
## [1] "vc:  0.435938454354556"
```

- Energy Intensity:

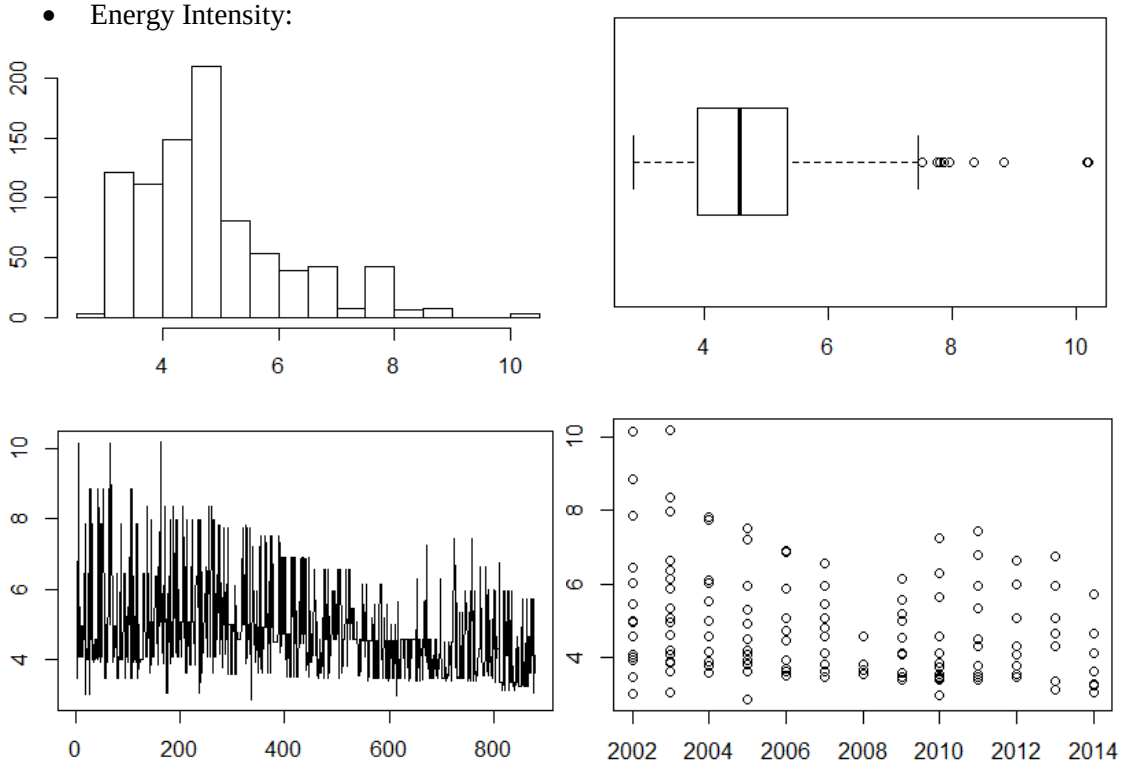


Figure 121

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.860   3.910   4.577   4.834   5.347  10.194
## [1] "sd:  1.31276154988043"
## [1] "vc:  0.271541754716896"
```

- Electricity Distribution Market:

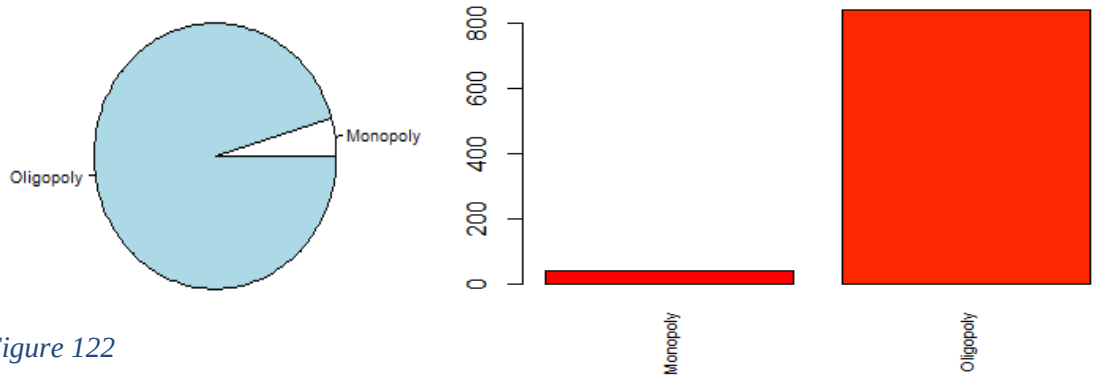


Figure 122

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  Monopoly Oligopoly
##        40       839
## [1] "Relative frequency table (proportions)"
##
```

```
##   Monopoly  Oligopoly
## 0.04550626 0.95449374
## [1] "Frequency table sorted"
## Oligopoly  Monopoly
##       839        40
## [1] "Relative frequency table (proportions) sorted"
##  Oligopoly   Monopoly
## 0.95449374 0.04550626
```

- Electricity Generation Market:

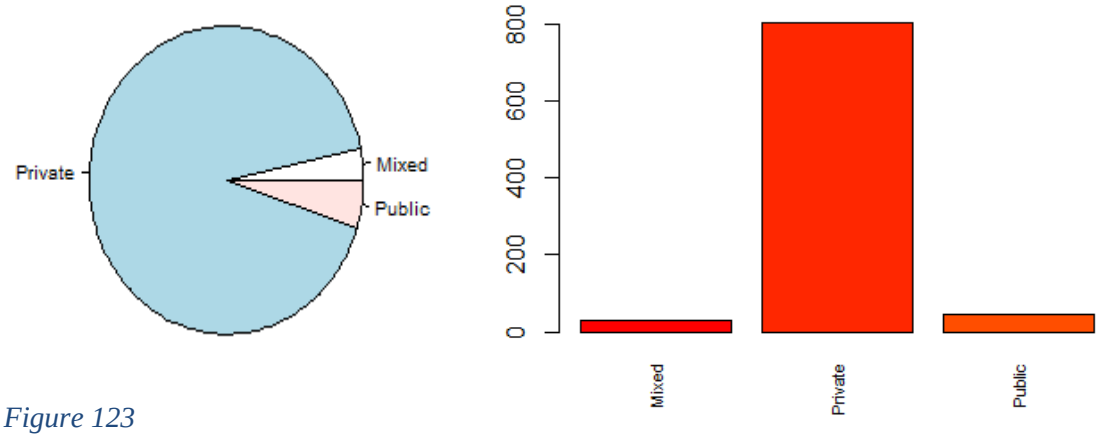


Figure 123

```
## [1] "Number of modalities:  3"
## [1] "Frequency table"
##   Mixed Private  Public
##      30     804      45
## [1] "Relative frequency table (proportions)"
##      Mixed    Private     Public
## 0.03412969 0.91467577 0.05119454
## [1] "Frequency table sorted"
## Private  Public   Mixed
##     804      45      30
## [1] "Relative frequency table (proportions) sorted"
##    Private     Public      Mixed
## 0.91467577 0.05119454 0.03412969
```

- Electricity Transmission Market:

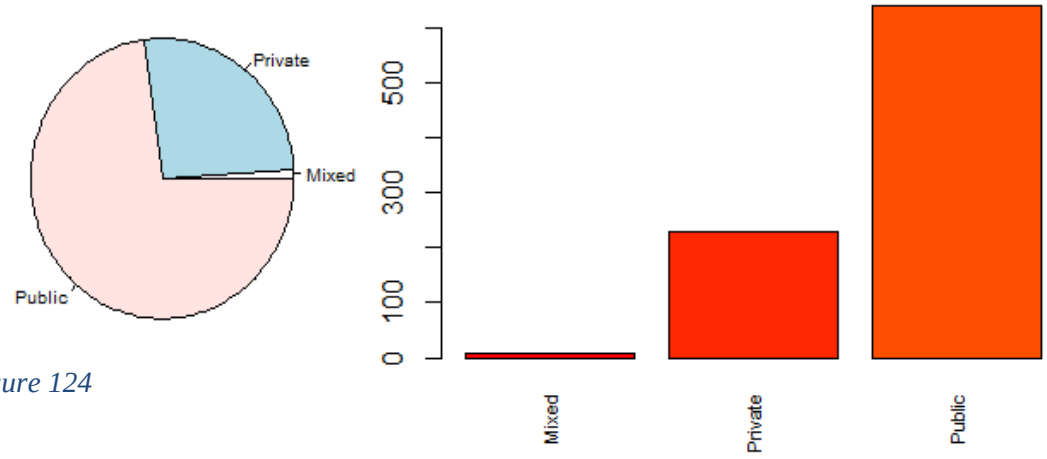


Figure 124

```
## [1] "Number of modalities:  3"
## [1] "Frequency table"
##   Mixed Private  Public
##       9     230     640
## [1] "Relative frequency table (proportions)"
##      Mixed    Private     Public
## 0.01023891 0.26166098 0.72810011
## [1] "Frequency table sorted"
##  Public Private   Mixed
##     640     230       9
## [1] "Relative frequency table (proportions) sorted"
##     Public    Private      Mixed
## 0.72810011 0.26166098 0.01023891
```

- Electricity Commercialization Market:

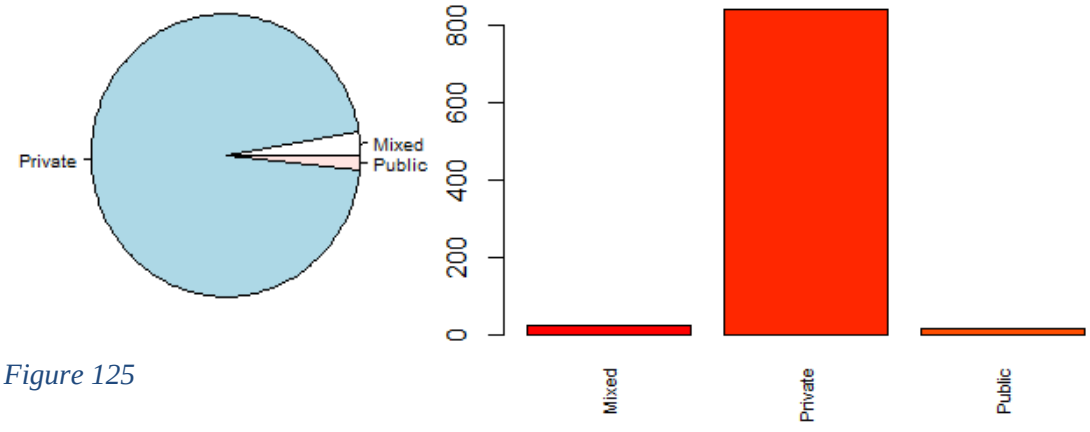


Figure 125

```
## [1] "Number of modalities:  3"
## [1] "Frequency table"
##   Mixed Private  Public
##      23     841      15
## [1] "Relative frequency table (proportions)"
##      Mixed    Private     Public
## 0.02616610 0.95676906 0.01706485
```

```
## [1] "Frequency table sorted"
## Private   Mixed  Public
##     841      23      15
## [1] "Relative frequency table (proportions) sorted"
##    Private      Mixed     Public
## 0.95676906 0.02616610 0.01706485
```

- Regulated Electricity Prices:

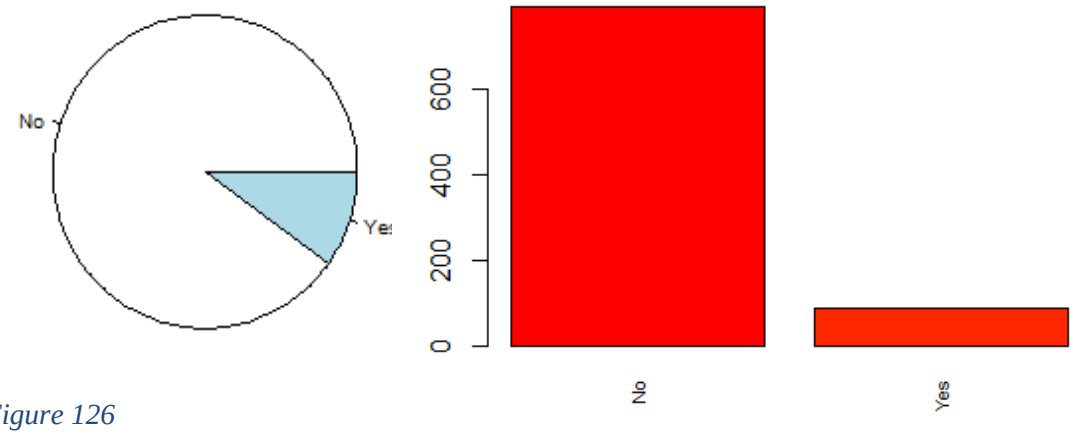


Figure 126

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
## 792  87
## [1] "Relative frequency table (proportions)"
##         No        Yes
## 0.90102389 0.09897611
## [1] "Frequency table sorted"
##
##  No Yes
## 792  87
## [1] "Relative frequency table (proportions) sorted"
##         No        Yes
## 0.90102389 0.09897611
```

- Interconnected Electric System:

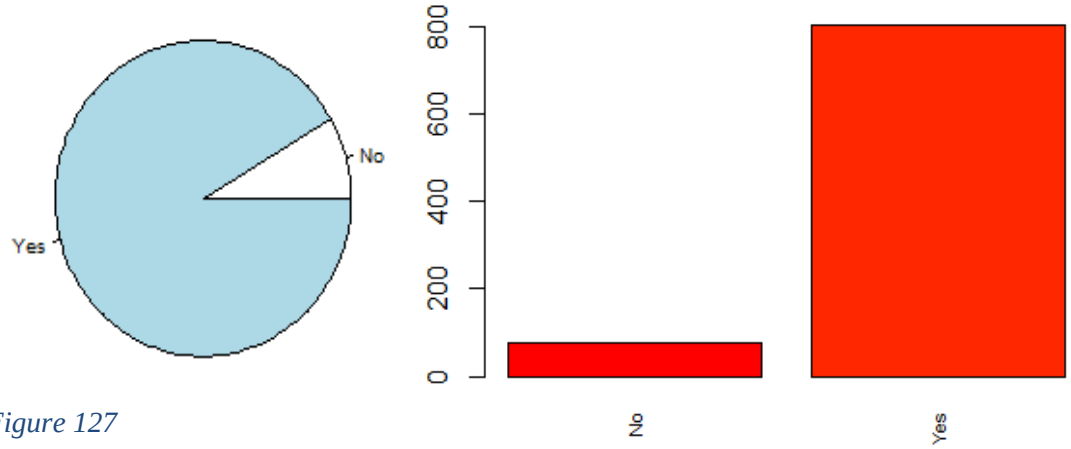


Figure 127

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
##  75 804
## [1] "Relative frequency table (proportions)"
##         No        Yes
## 0.08532423 0.91467577
## [1] "Frequency table sorted"
## Yes  No
## 804  75
## [1] "Relative frequency table (proportions) sorted"
##        Yes         No
## 0.91467577 0.08532423
## [1] "variable  Nuclear : Nuclear"
```

- Nuclear Plants:

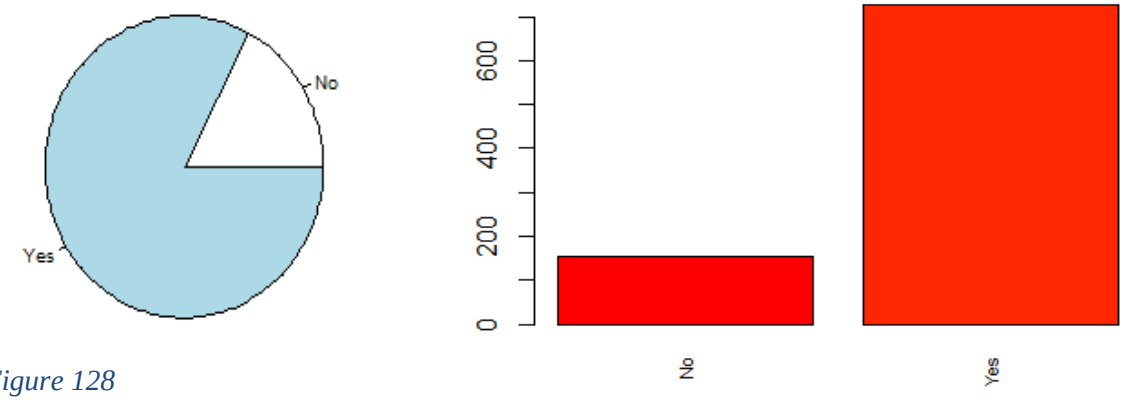


Figure 128

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
## 153 726
## [1] "Relative frequency table (proportions)"
##        No       Yes
```

```
## 0.1740614 0.8259386
## [1] "Frequency table sorted"
## Yes  No
## 726 153
## [1] "Relative frequency table (proportions) sorted"
##       Yes        No
## 0.8259386 0.1740614
## [1] "variable  RatParis : RatParis"
```

- Ratification of the Paris Agreement:

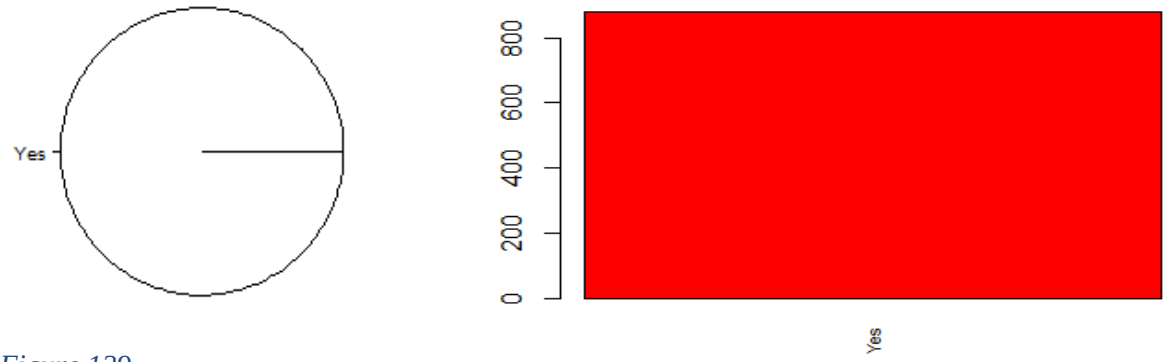


Figure 129

```
## [1] "Number of modalities:  1"
## [1] "Frequency table"
## Yes
## 879
## [1] "Relative frequency table (proportions)"
## Yes
##   1
## [1] "Frequency table sorted"
## Yes
## 879
## [1] "Relative frequency table (proportions) sorted"
## Yes
##   1
```

- Number of electric substations in the country:

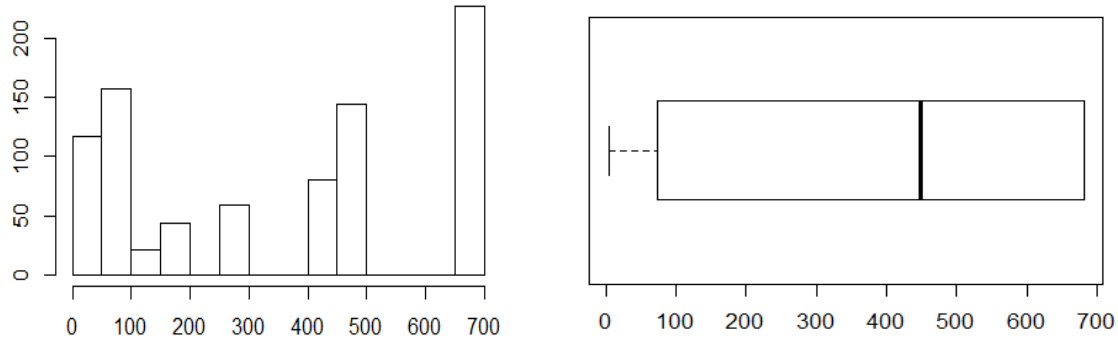


Figure 130

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     4.0    74.0   449.0   352.8   681.0   681.0      30
## [1] "sd:  254.347960944832"
## [1] "vc:  0.720960669747702"
```

- Number of Blackouts per year:

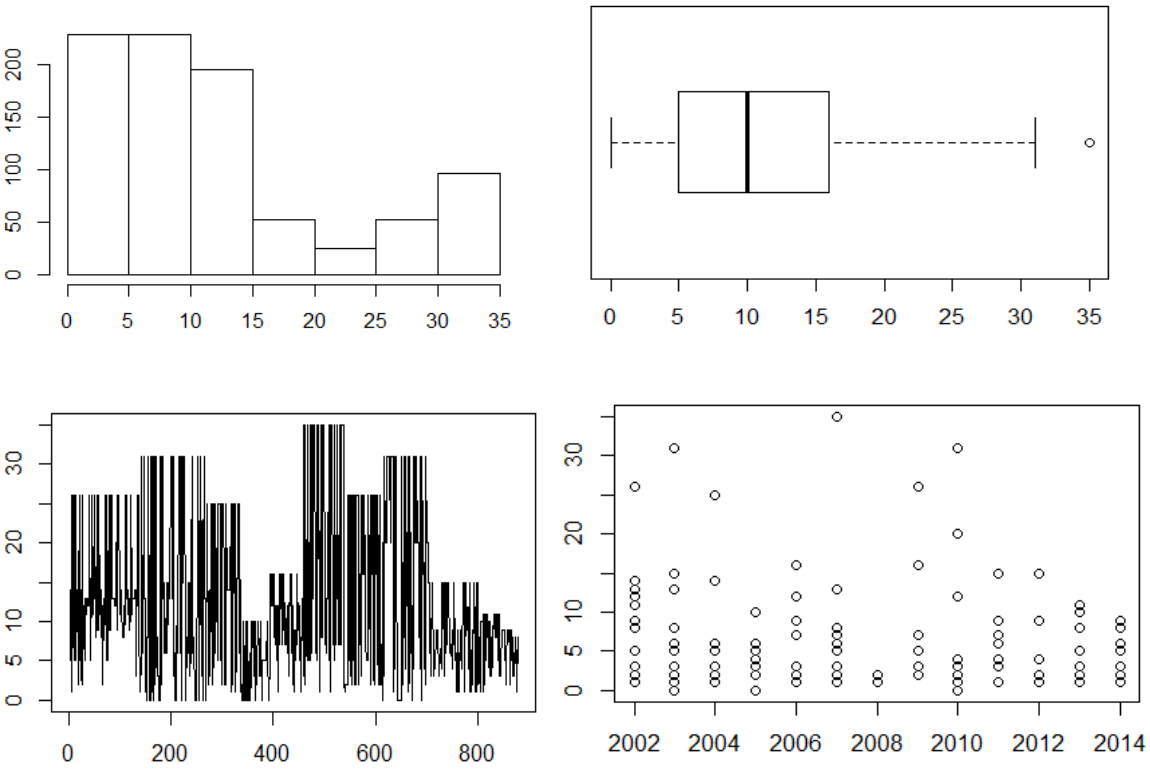


Figure 131

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    5.00   10.00   12.48   16.00   35.00
## [1] "sd:  9.73400353072277"
## [1] "vc:  0.780247045732748"
```

Climate variables:

- Type of Climate:

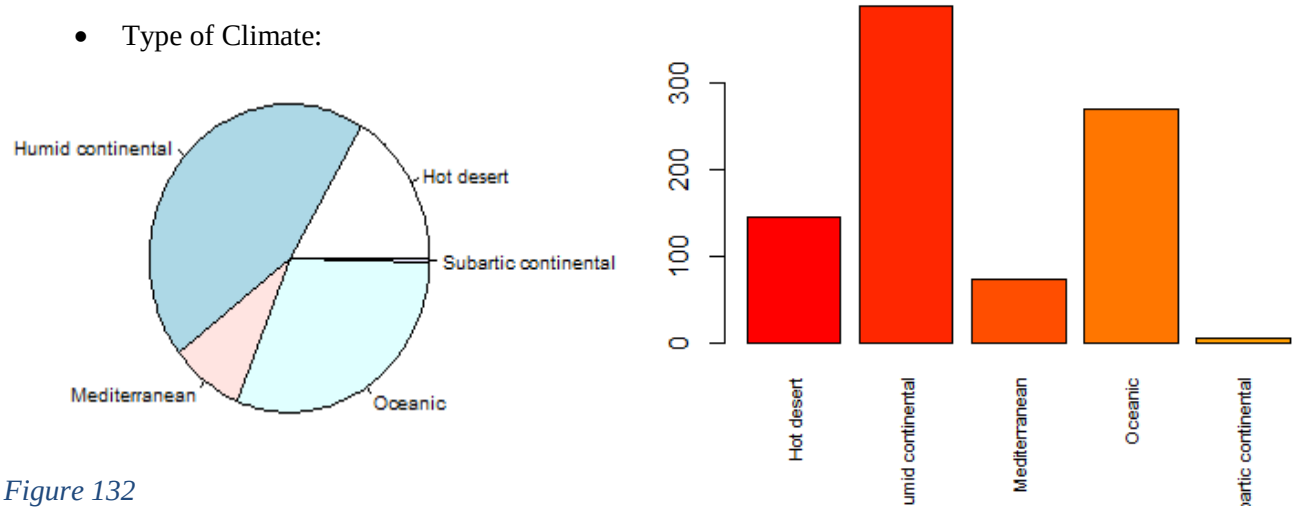


Figure 132

```
## [1] "Number of modalities:  5"
## [1] "Frequency table"
##          Hot desert    Humid continental       Mediterranean
##                 144                 388                  73
##             Oceanic Subartic continental
##                 270                   4
## [1] "Relative frequency table (proportions)"
##          Hot desert    Humid continental       Mediterranean
##         0.163822526         0.441410694         0.083048919
##             Oceanic Subartic continental
##         0.307167235         0.004550626
## [1] "Frequency table sorted"
##    Humid continental             Oceanic          Hot desert
##                  388                 270                 144
##       Mediterranean Subartic continental
##                   73                   4
## [1] "Relative frequency table (proportions) sorted"
##    Humid continental             Oceanic          Hot desert
##         0.441410694         0.307167235         0.163822526
##       Mediterranean Subartic continental
##         0.083048919         0.004550626
```

- Island:

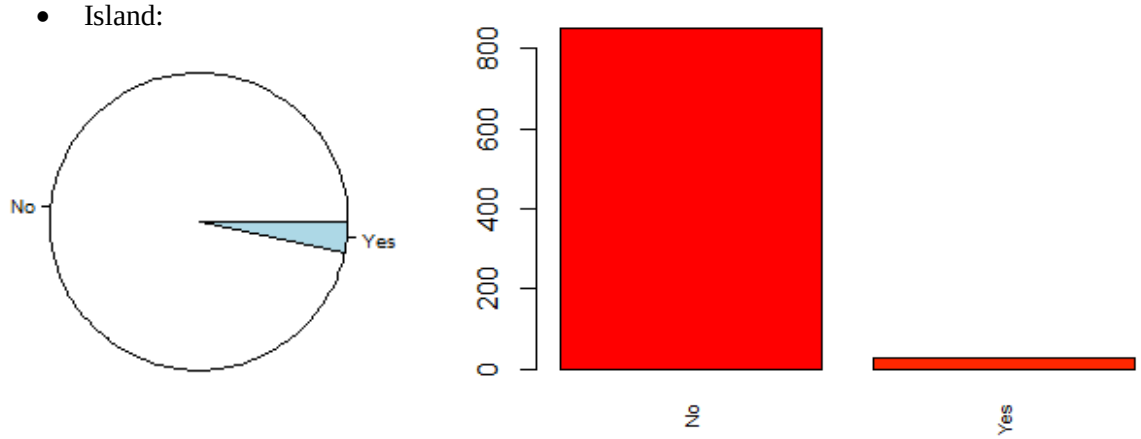


Figure 133

```
## [1] "Number of modalities:  2"
## [1] "Frequency table"
##  No Yes
## 850  29
## [1] "Relative frequency table (proportions)"
##         No        Yes
## 0.96700796 0.03299204
## [1] "Frequency table sorted"
##  No Yes
## 850  29
## [1] "Relative frequency table (proportions) sorted"
##         No        Yes
## 0.96700796 0.03299204
```

- Average Temperature:

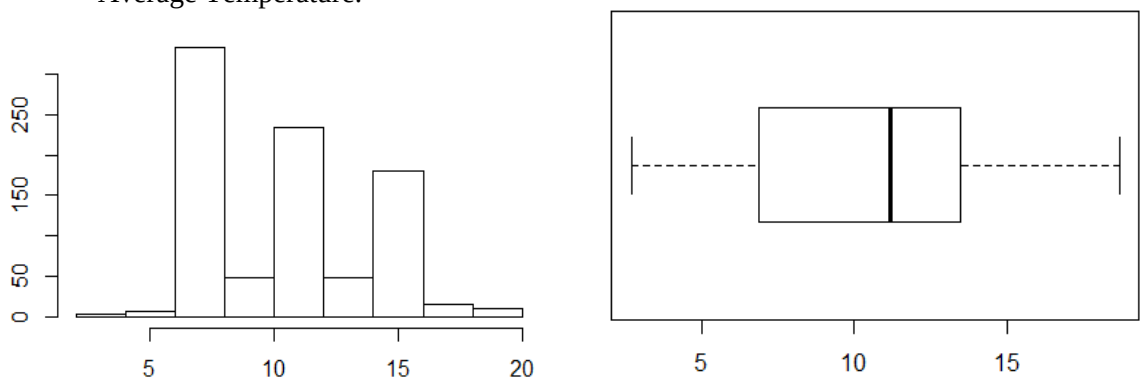


Figure 134

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.
Max.
##    2.70    6.90   11.20   10.61   13.50   18.70
## [1] "sd:  3.54980758981488"
## [1] "vc:  0.334689943198713"
```

- Average Precipitation:

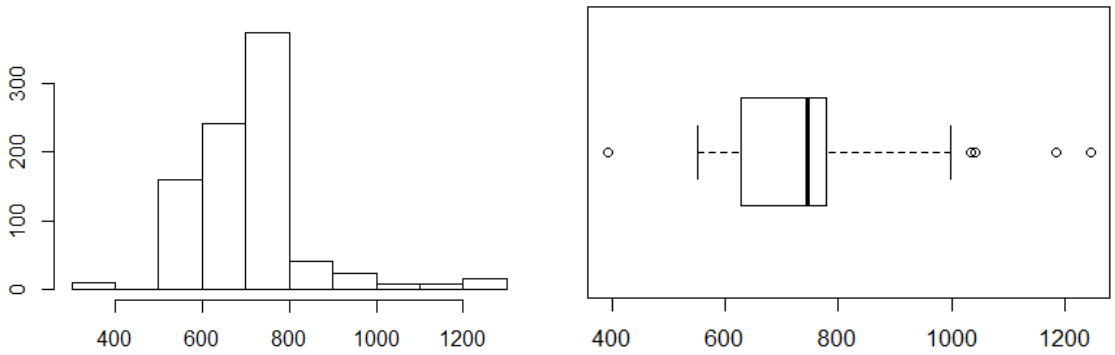


Figure 135

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   391.1   626.2   745.6   719.4   779.5  1245.8
## [1] "sd:  135.939932758367"
## [1] "vc:  0.188968101092261"
```

8.3 Appendix C. Basic descriptive of the variables regarding the clustering profiling

We use the same division of the variables we have been using along the document, in order to follow the logics and dynamics of the whole study.

Economic variables:

- GDP per Capita:

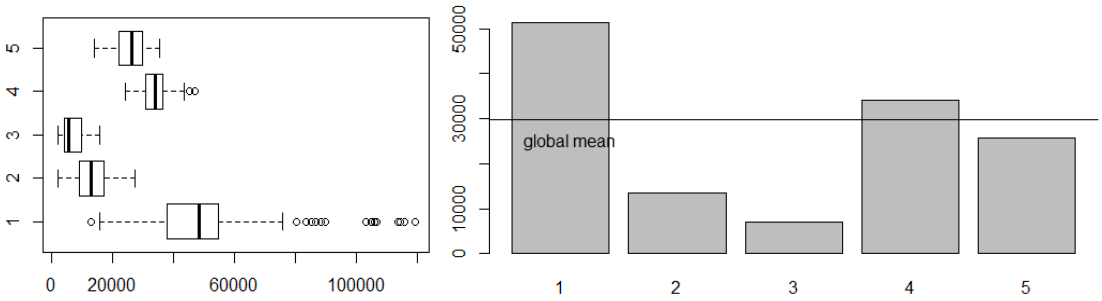


Figure 136

As we can see in both graphics, the boxplot and the barplot, the first group has the bigger levels of GDP per capita, in terms of mean but also of variability between countries. Followed by the fourth and the fifth group, which are very similar among them. The second and third groups have big differences in terms of GDP per capita with the rest of groups, having smaller mean values of it, but also, being more compact in their behavior as a group.

- GINI Index:

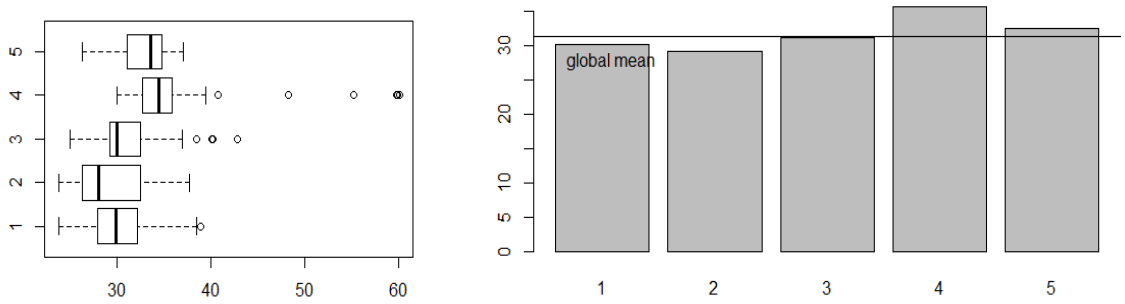


Figure 137

The GINI Index is not a discriminant variable that provides with big differences from one group to another. All of their mean values are quite similar, being the fourth one the better positioned. The variance of the values between countries in the different groups is also very similar, but for the appearance of some superior outliers in the fourth group.

- Total Population:

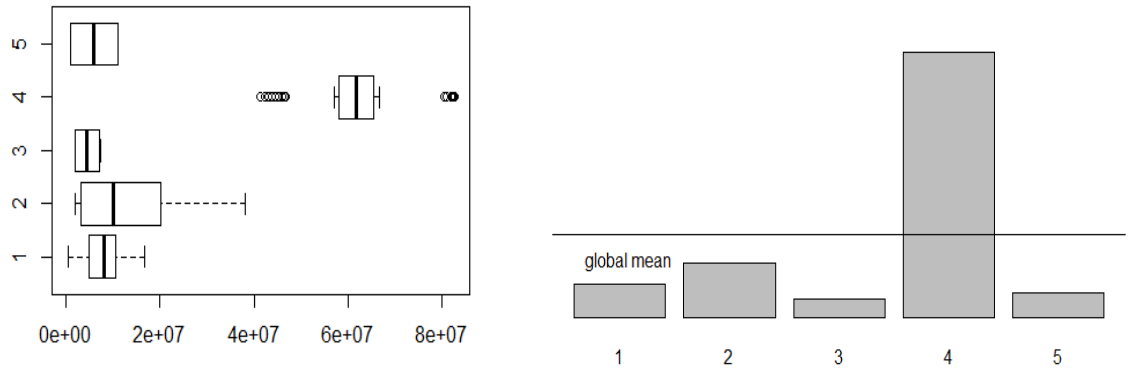


Figure 138

The fourth group counts on a total population considerably bigger than the rest in terms of mean values. The variable is quite similar in each of the countries forming the group even if there are some superior and inferior outliers on it. The rest of groups have quite smaller populations, being the second bigger one, but far away from the levels of the fourth group.

- Rural Population:

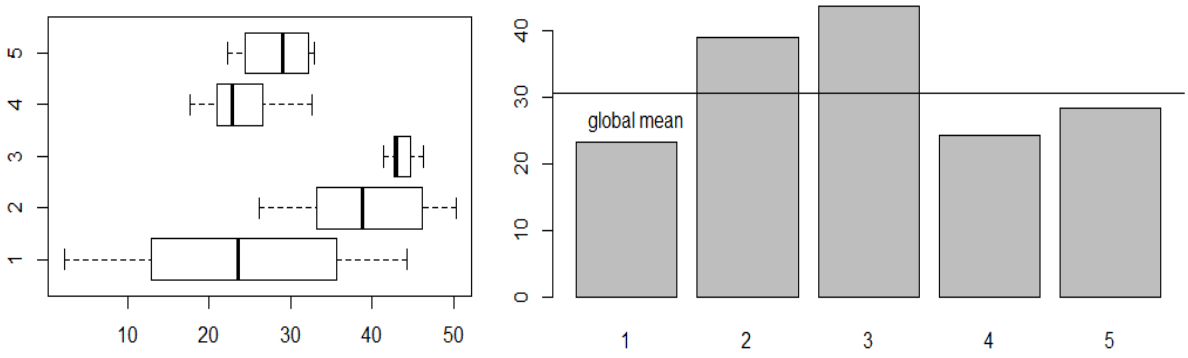


Figure 139

The graphics show that the third and second groups have the bigger levels of rural population while the rural population in the fifth, fourth and first group are smaller and in the same scale of values. It is also important to say that, the variance of the first group is the most important one, so even if the mean value is low, there will be countries in it with bigger rural population, at the same level of the second and third groups.

- Urban Population

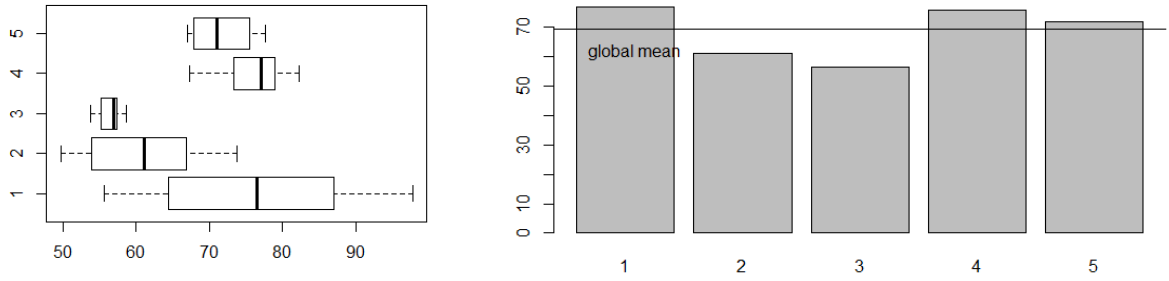


Figure 140

The urban population is the complementary variable of the rural population. Consequently, the first, fourth and fifth group have bigger values of this kind of population and the second and third, smaller ones. Again, the first group show important differences between the countries composing the group, as we can realize due to its big variance.

- Corruption Ranking

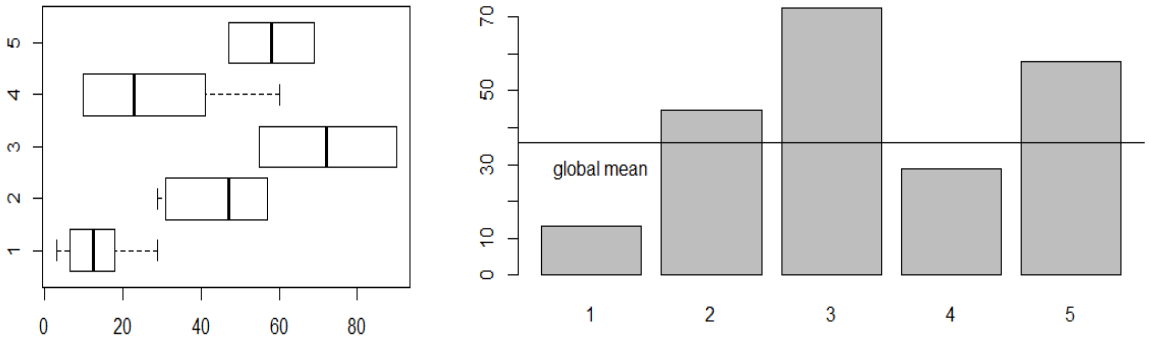


Figure 141

This variable looks as a discriminant one in the clustering creation as it has different behaviors for every group of countries. The third ones seems to be the most corrupted one followed by the fifth and second group. Group number four holds better positions in this ranking and the first one is the best positioned. In terms of variability all the groups behave similar but the first one, which is more compact.

- Democracy Ranking:

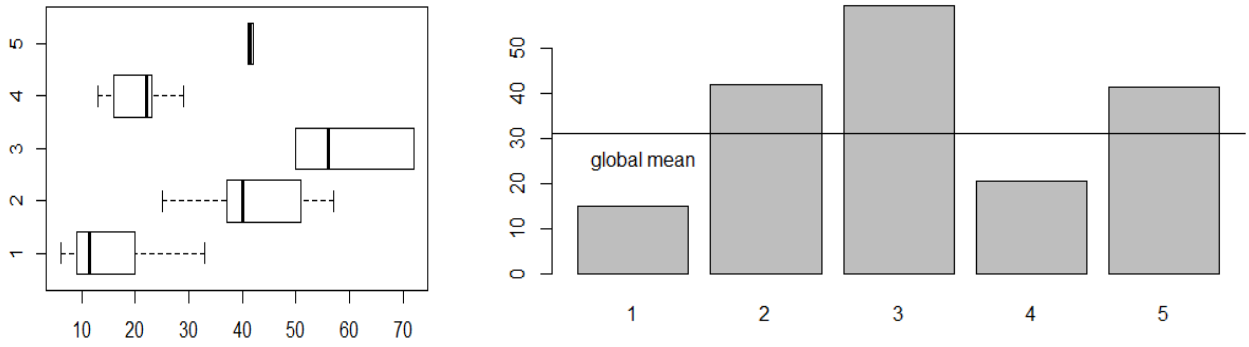
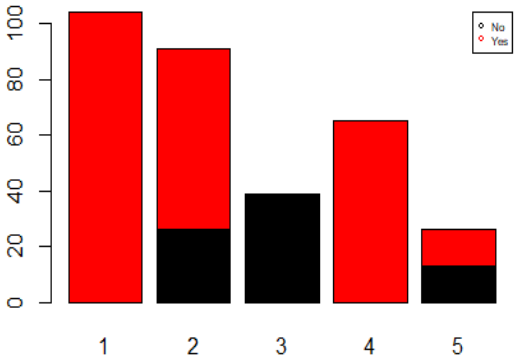


Figure 142

It is important to highlight that, the highest levels of the variable Democracy Ranking, means lower positions on the ranking. That is to say, if you have a value of 30 you are worse positioned on the ranking than a country with a value of 10. Therefore, the second and third group are the worst positioned on the democracy ranking, followed by the fifth group. The fourth and first group hold better positions in this ranking. The variability between countries in the same group is similar, except for the fifth group, which behaves in the same way.

- Belonging to the OCDE:



While the first and fourth group are fully composed by countries belonging to the OCDE, the third group has no country in this organization. In the second group almost all countries are in the OCDE but for Romania and Lithuania. In the fifth group Cyprus does not belong to the OCDE.

Figure 143

- Type of Government:

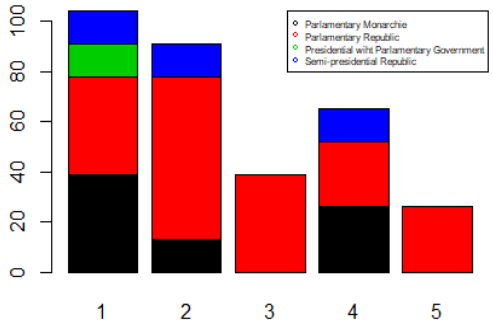


Figure 144

The type of country is not a discriminant variable when clustering the countries, as for most of the groups, there are no coincidences about the type of government among the members of the groups. There is the exception of the third and fifth groups, which only contain countries with Parlamentary Republics.

- Belonging to the European Union:

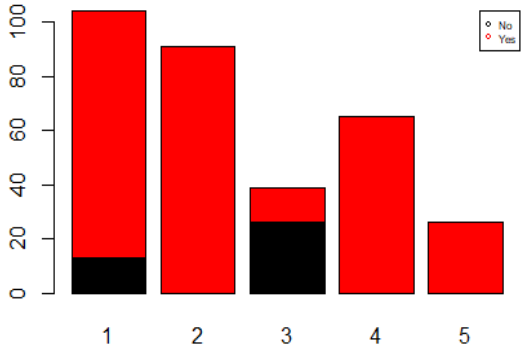


Figure 145

There are only a few countries outside of the European Union (Switzerland in the first group and Serbia and Macedonia in the third one). Almost all of the countries belong to this organization so it does not look as an important value when separating countries into different groups.

Energy Variables:

- Emission of CO_2:

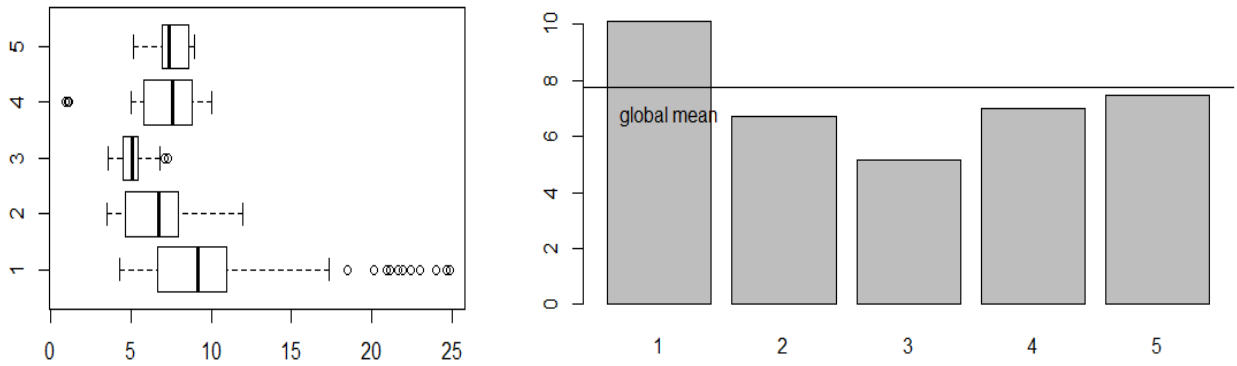


Figure 146

Even if, in absolute values, the CO_2 emissions of all the groups are not very different, there is the first group, which has the biggest mean value of emissions as well as the biggest variance among the countries composing it. The other groups have likely levels of emissions and variances, being the third group the most compact and less pollutant in terms of CO_2.

- Access to Electricity in terms of Total Population:

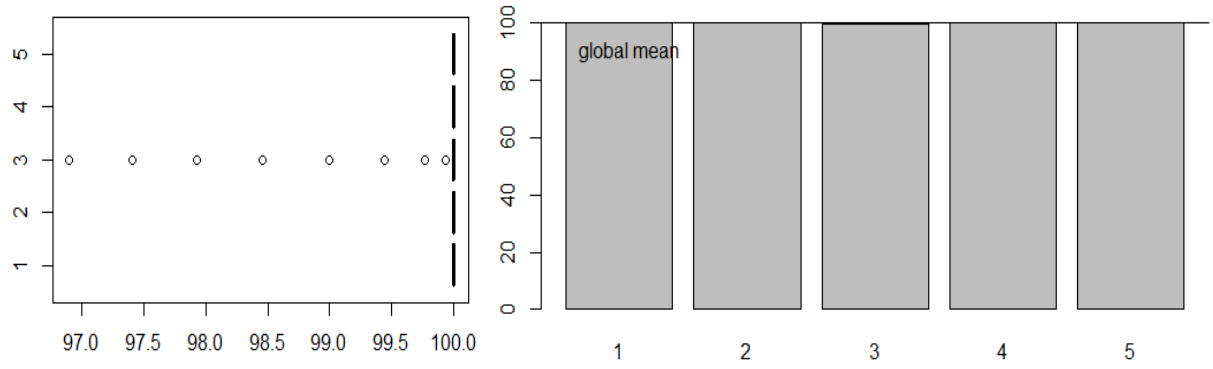


Figure 147

As we can see in the graphics, the access to electricity in terms of total population in the countries shows no difference between groups. Even if the third group has some inferior outliers, considering the absolute values of the variable, it is not a discriminant variable.

- Access to Electricity in terms of Rural Population:

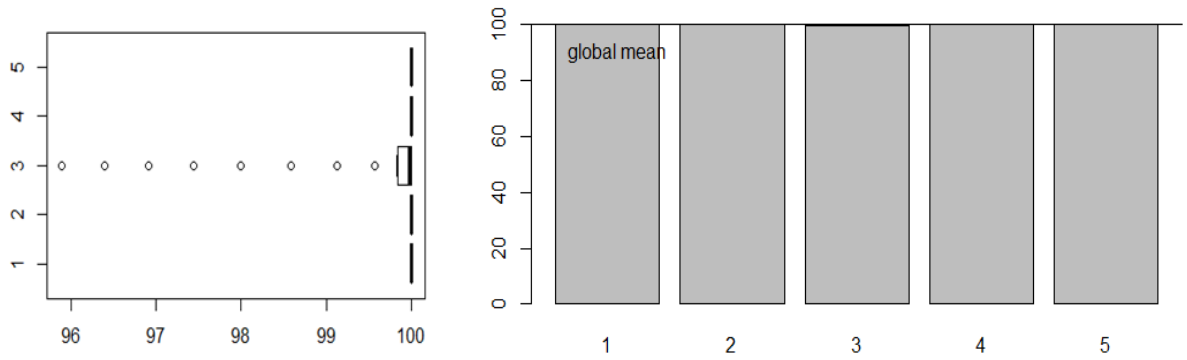


Figure 148

As it happened with the last variable, the access to electricity of the rural population shows any significant difference between the groups and is not critical when composing the groups.

- Electricity Production from Hydrological Sources:

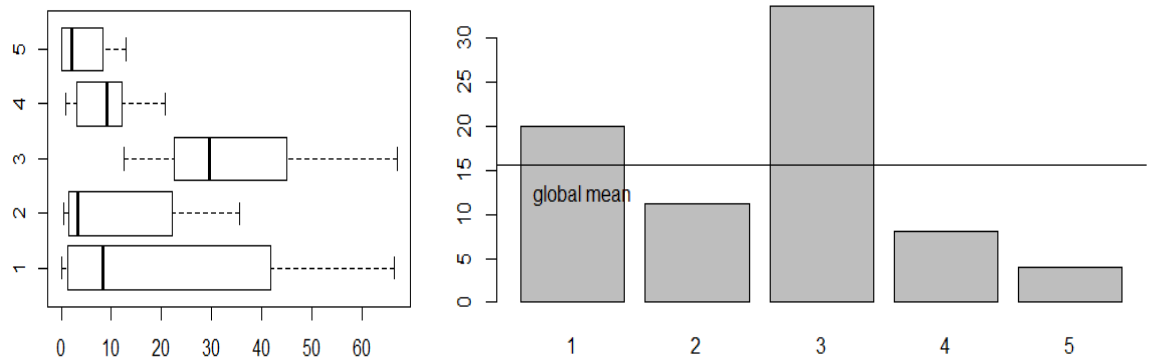


Figure 149

We can perceive important differences between groups regarding to their electric production from hydrological sources. The third group is the one with the most important hydroelectric production followed by the first group. Second, fourth and fifth group show smaller amounts of electricity produced from this source. Talking about the variability between groups, the first and third ones are also the more heterogeneous.

- Electricity Produced from Nuclear Sources:

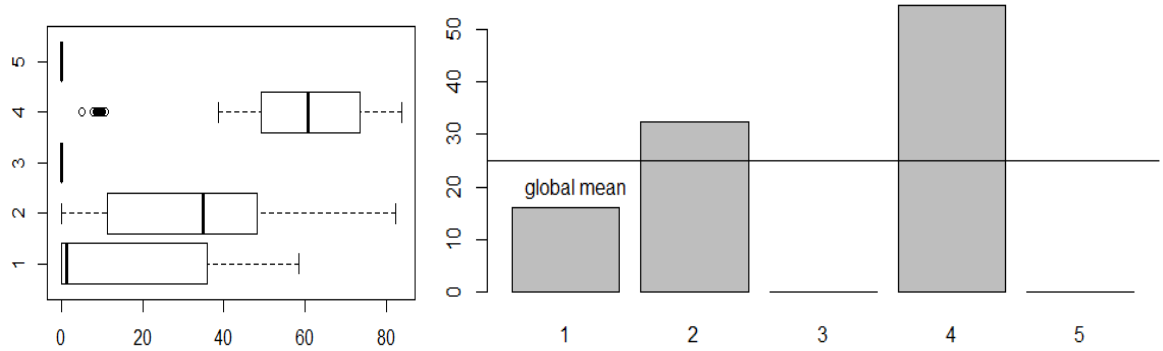


Figure 150

This variable shows two opposite behaviours in the groups: those producing electricity from nuclear plants (1, 2 and 4) and those which slightly use this source of energy (3 and 5). The group producing more electricity from nuclear plants is the fourth. It was to be expected as Germany and France, countries with a long tradition in nuclear production, belong to this group. The second group and the first produce less nuclear electricity but they are less homogeneous as a group.

- Electricity Production from Oil, Gas and Coal Sources:

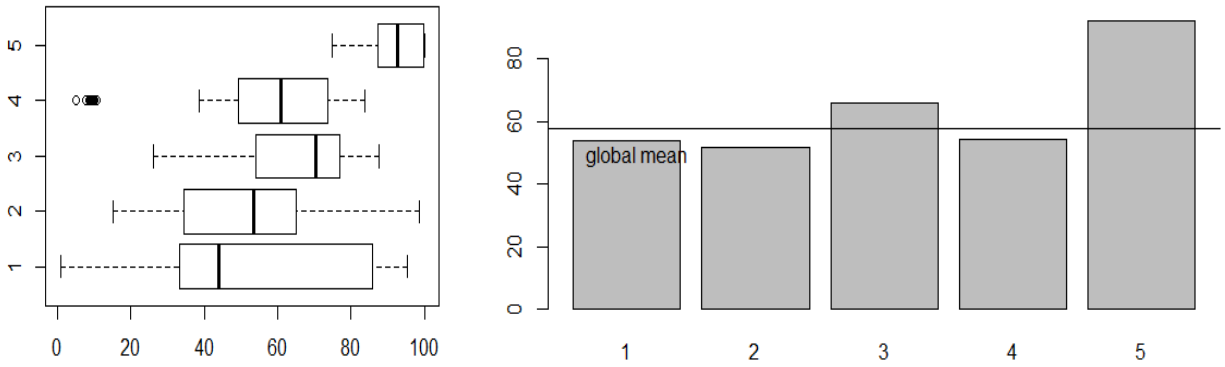


Figure 151

All of the groups are in the same range of electricity production from fossil fuels. Nevertheless, the fifth group shows the higher values and lower variance between countries, separating itself from the others. The rest of the behave similarly, even if the third group has bigger mean values and the first one is more heterogeneous in terms of this kind of electricity production.

- Electricity Produced from Renewable Sources:

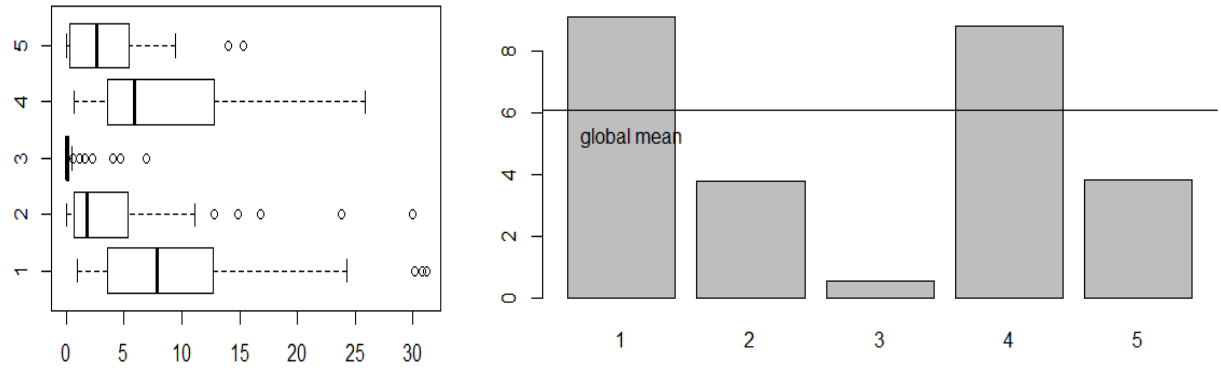


Figure 152

While the first and fourth groups produce bigger amount of electricity from renewable sources the third one produce very little. The second and fifth group produce some but in lower levels than the first and fourth. It is remarkable that, even if those two groups lead the renewable production the behavior of the countries composing them is not homogeneous. The electricity production from clean sources looks like a discriminant variable that helps to model the profiles of the different clusters.

- Energy Imports:

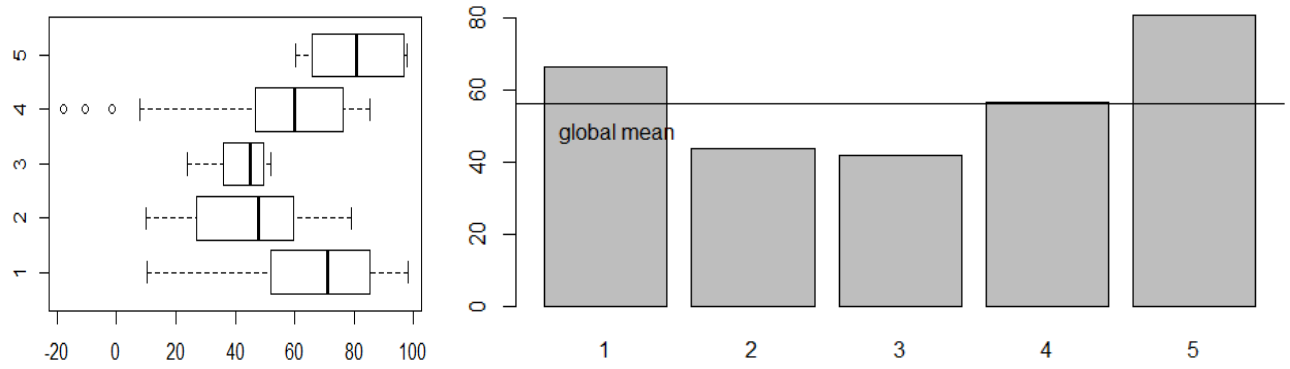


Figure 153

All the groups count on relative high levels of imported energy. Nevertheless, the fifth group, followed by the first are the ones with bigger amounts of energy imported. The difference is that the fifth group has a more homogeneous behaviour. The fourth, second and third group import less energy, being the third one the less demanding in this sense; and the fourth one the most variable in the imports of the countries in it.

- Electric Power Consumed:

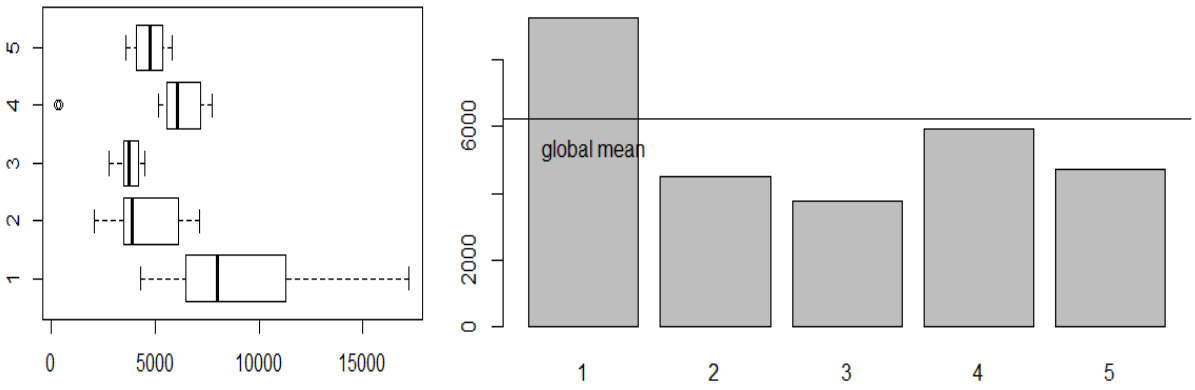


Figure 154

The electric power consumption differs strongly from one group to another. The first one is the most consuming group, although there are differences among the countries composing it. The third, second and fifth one consume considerably less electric power, and the fourth group shows medium levels. These four groups are much more compact in their consume tendencies as a group.

- Electric Power Transmission and Distribution Losses.

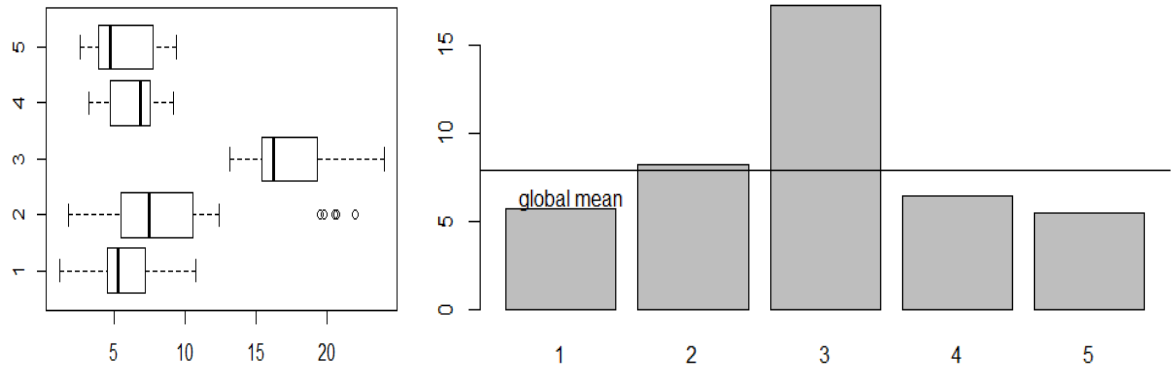


Figure 155

The losses that the power transmission and distribution cause are low and similar for the first, second, fourth and fifth group, having close mean levels and variability. Nevertheless, the third group has more losses than any other country, showing a particular feature of this group.

- Fuel Exports:

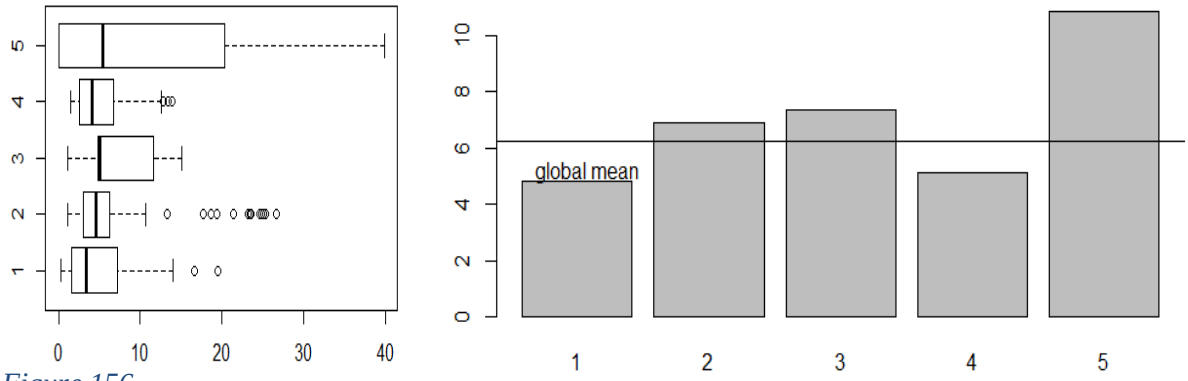


Figure 156

The fifth group, being the one importing more energy, is also the group that exports more fuel products. The variability is caused by annual differences on the exports. The other groups export in a similar way in average, being the mean value of the third and second group slightly bigger (on the second group due to some superior outliers).

- Energy Intensity:

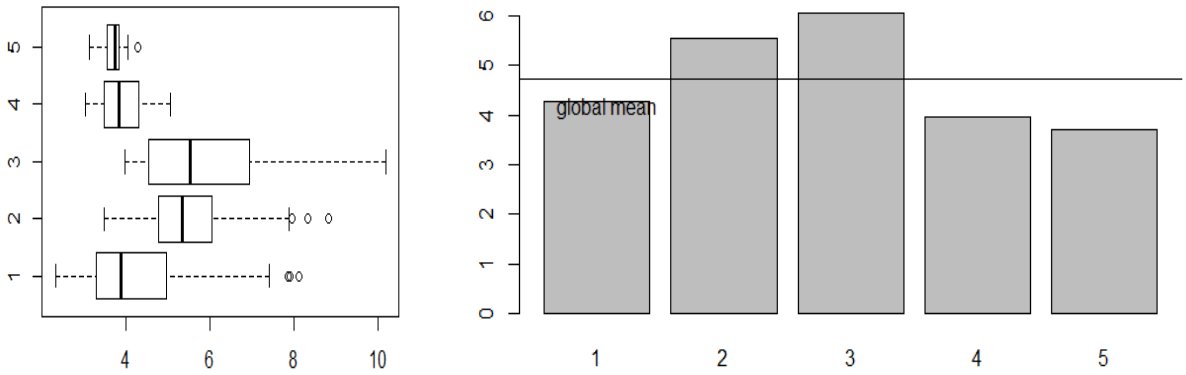
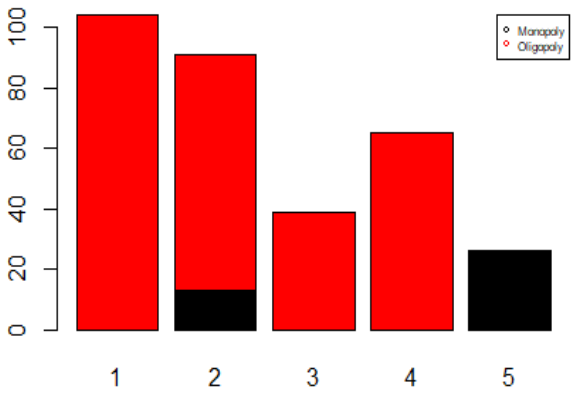


Figure 157

The Energy Intensity, the variable that shows, somehow, the energy efficiency of a country as a whole, is bigger in the third group, followed by the second one. The first group presents big variances between country causing the growth of the mean value while the fifth and four group have the smaller energy intensities. The variances of the groups are unalike, the first, second and third ones have a lot of dissimilarities between their countries while the fourth and fifth are more homogeneous.

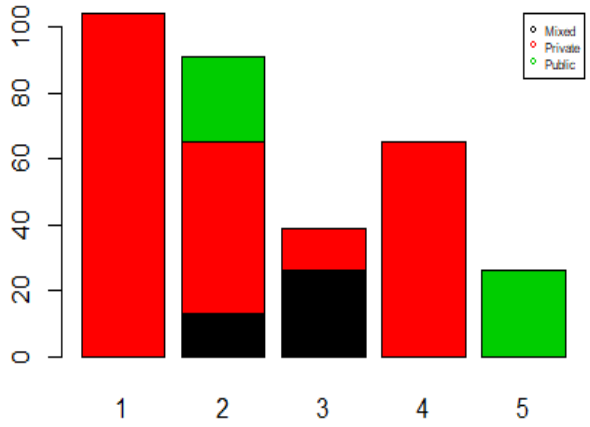
- Electricity Distribution Market:



In general, all the countries count on oligopolistic markets regarding the electricity distribution. The exception are Greece and Cyprus (forming the fifth group) and Slovenia (in the second group) which have monopolistic markets.

Figure 158

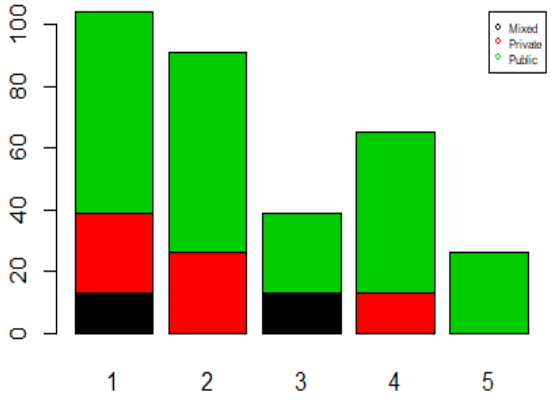
- Electricity Generation Market:



The European countries mostly have private generation markets (the first and fourth group are completely private) but there are some cases of mixed systems (Croatia and Macedonia in the third group and Lithuania in the second) and of public markets (Slovenia and Hungary in the second group and the fifth group as a unity). Despite of having some groups completely homogeneous on its generation market others are not, what makes this variable look as a not determinant one.

Figure 159

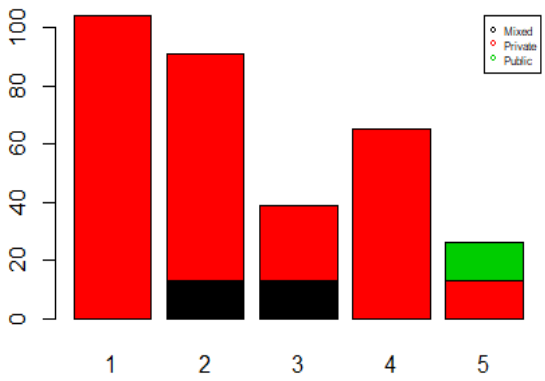
- Electricity Transmission Market:



Public transmission markets are majority in every group. There are few examples of private systems like Switzerland and Finland (1st group), Romania and Czech Republic (2nd group) and Germany (4th group). In addition, there exist some examples of mixed systems like Belgium in the first group and Croatian in the third one.

Figure 160

- Electricity Commercialization Market:



As it happened with the generation market, in general, the commercialization of the electricity depend on private enterprises in Europe. There are some exceptions as Lithuania (2nd group) and Serbia (3rd group) with mixed commercialization markets and Greece with the sole public market (5th group).

Figure 161

- Regulated Electricity Prices:

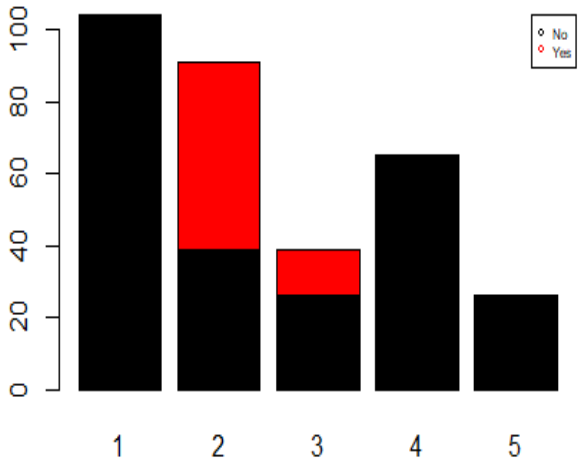


Figure 162

As a consequence of the European energy legislation, all the countries present in the study either have liberalized prices of the electricity or are in transition process to establish them. The few countries that still count on regulated prices for their electricity are Slovenia, Polonia, Hungary and Lithuania in the second group and Serbia in the third one.

- Interconnected Electric System:

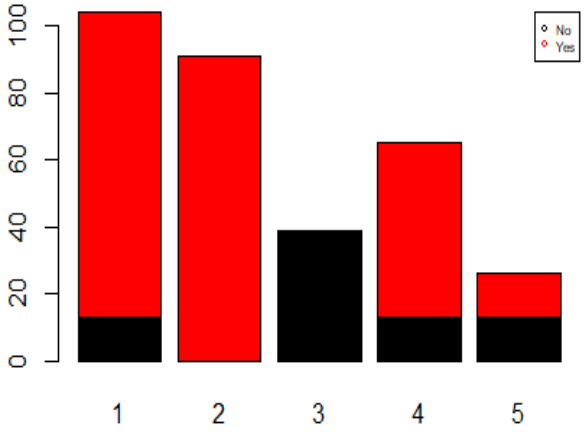


Figure 163

There are six regional wholesale electricity markets in Europe, accordingly, a lot of the countries we are studying belong to these six markets. Even though, there are some countries that do not belong to one of them and still produce and consume all their electricity inside their frontiers. These countries are Luxembourg (1st group), Serbia, Macedonia and Croatia (3rd group), Italy (4th group) and Cyprus (5th group).

- Nuclear Plants:

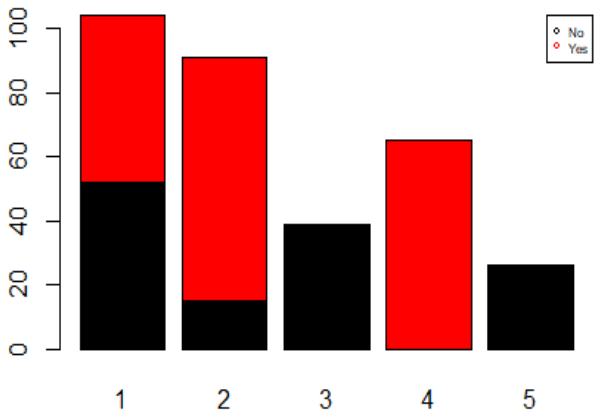
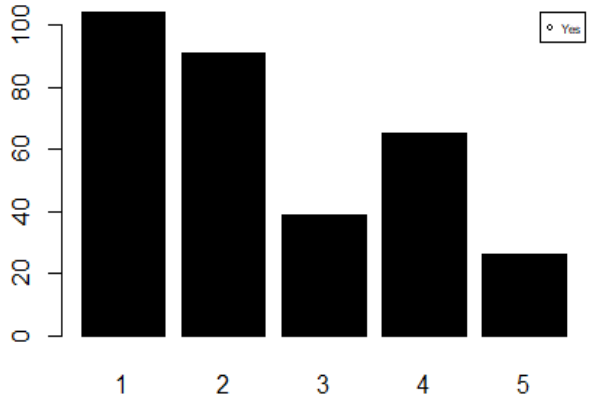


Figure 164

This variable complements the information given by the variable *ElecProdNuc*. Countries not producing electricity from nuclear fuels are, in general, those, which do not count on nuclear plants in their territory. The exception are those countries that do not have nuclear plants but buy this kind of electricity from other countries. The countries that do not have nuclear plants

are Ireland, Portugal, Luxembourg and Austria from the first group, Polonia form the second. The third and fifth group are completely free of nuclear plants.

- Ratification of the Paris Agreement:



All the countries on the dataset have singed the Paris Agreement so this variable does not implies any dissimilarity between groups.

Figure 165

- Number of electric substations normalized by the population of the country:

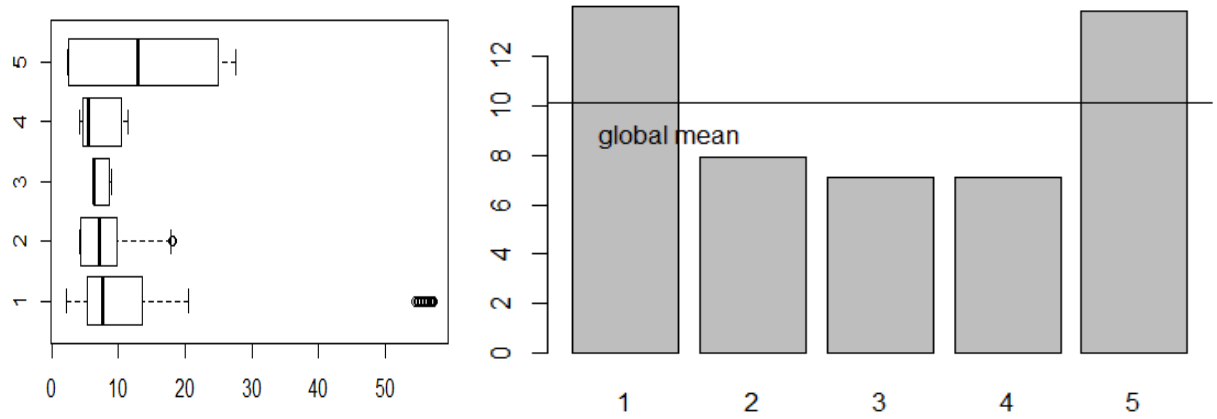


Figure 166

The variable number of electric substation is more useful if we contextualize it with the total population of every country as we explained in the Pre-process section. Using the transformation, we realize that the countries with more number of substations per capita are the first and fifth group (even if the fourth group is the one with more substations in absolute values). The fifth group counts on the bigger mean values while the mean value of the first group increases due to the superior outliers. The second, third and fourth group have a moderated number of electric substations in comparison with the others.

- Number of blackouts per year normalized by the population of the country:

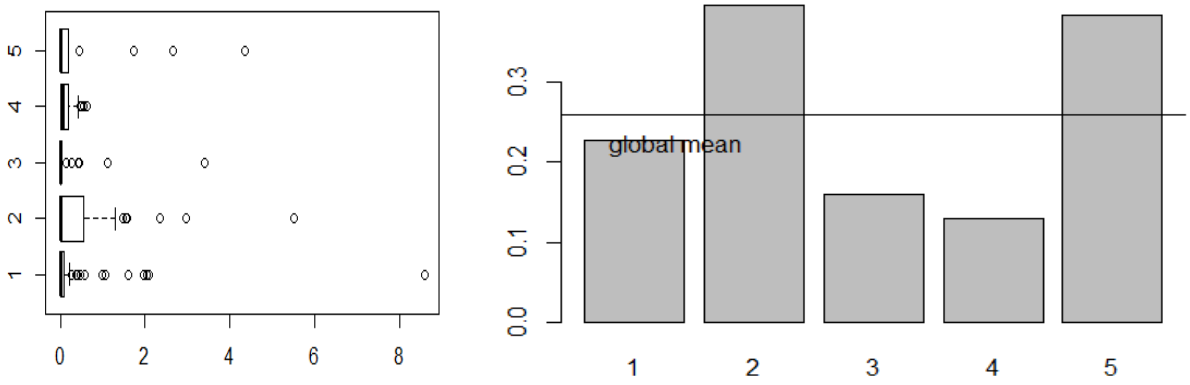


Figure 167

Following the same procedure with the variable *BoperYear* we obtain the number of blackouts during a year in a country for every 1 million of people. Although the group with more failures during a year is the fourth, when we normalize the value it is the one showing the lowest number of blackouts. On the contrary, the second and fifth group bear with the higher number of failures during a year. Nonetheless, it is important to highlight that the differences in the mean values are due to the outliers present in the groups, caused by punctual failures in the system. In general, the number of incidences in all the groups follow the same tendencies.

- Number of blackouts per year normalized by the number of electric substations of the country:

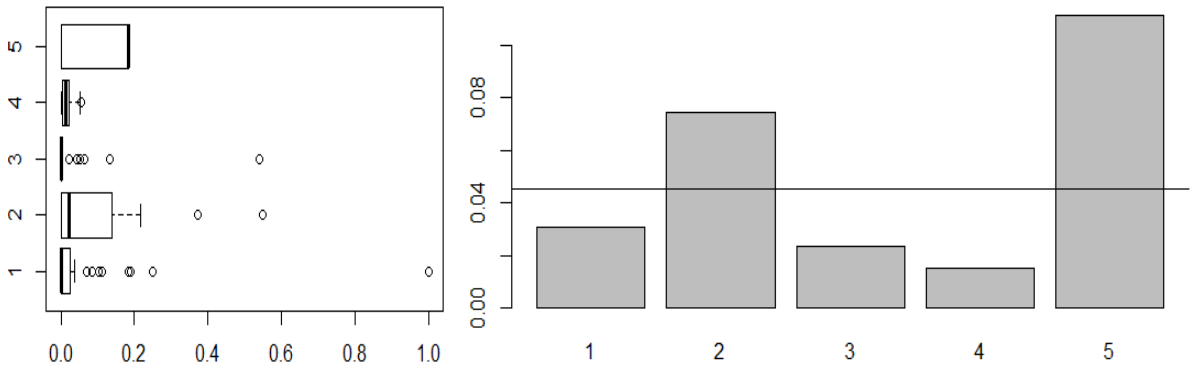


Figure 168

Another way to normalize the variable *BoperYear* is to divide it by the number of electric substations of the country (*Nsub)*. We find again that the fifth and second groups are the ones with more number of failures on the system. The first, third and fourth groups present less failures per substations. It is remarkable that the absolute scale of the variable small.

- Average Energy Not Supplied during the Failure:

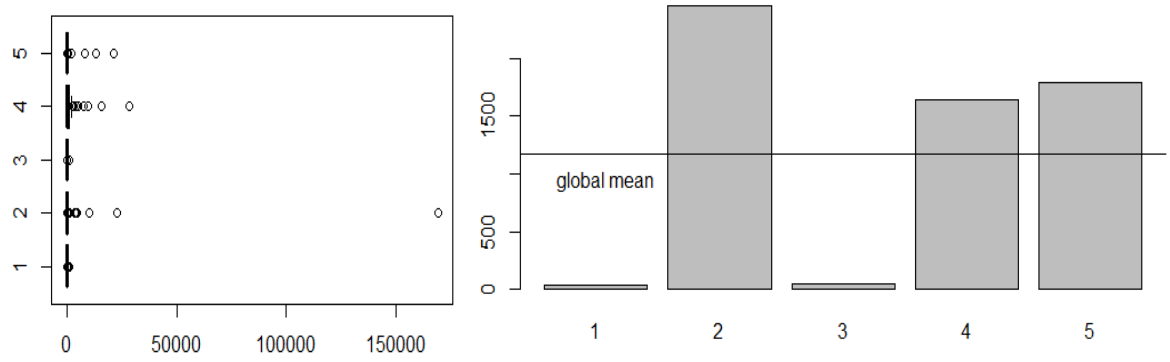


Figure 169

The absolute scale of this variable is very small, so even big differences between groups does not mean real big differences perceived by the users. Moreover, as it was explain in the Pre-process section, calculating a mean value of a variable with a heavy-tailed distribution is not the best statistical analysis that we can do, so values are just simple and quick approximations to understand the behavior of the countries.

Nevertheless, analyzing the results we can see that the second and fifth group suffer from more important failures in terms of energy not supplied, followed by the fourth group, while the first and third one show more static values of energy not supplied, around zero.

- Average Equivalent Time of the Blackout Taking into Account the Amount of Service during 12 Months:

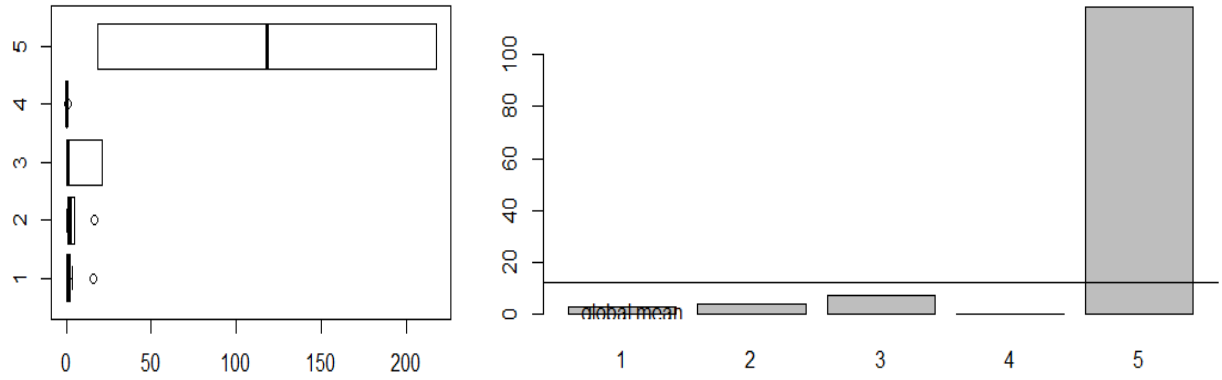


Figure 170

The same considerations than before must be taken when dealing with this variable. Assuming that the behavior of the groups regarding to this variable is just a guidance, we realize that the fifth group present the bigger times of interruption of the electric service due to failures. The other groups are all around zero.

- Maximum Equivalent Time of the Blackout Taking into Account the Amount of Service during 12 Months:

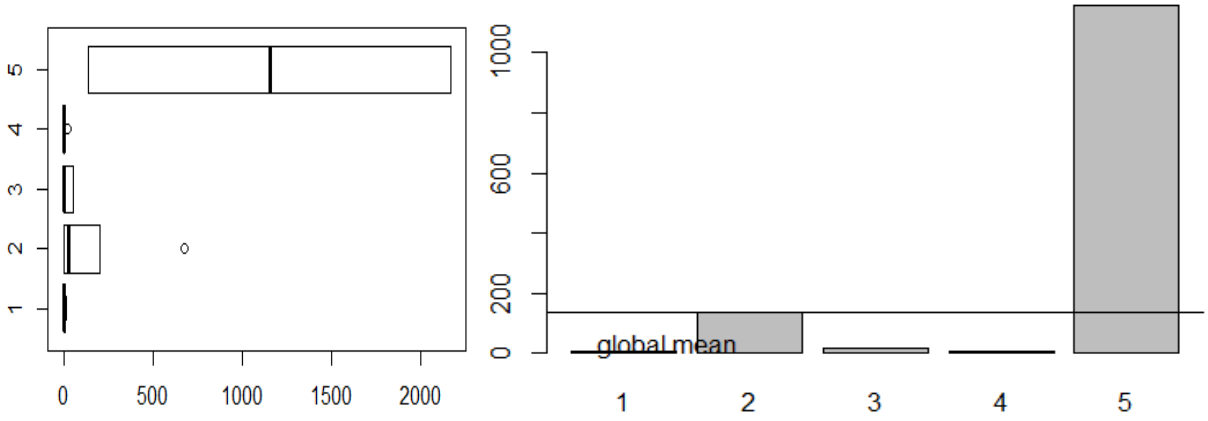


Figure 171

With the aim of complementing the last variable we study the interval the equivalent time of interruption has in every group. Again, the group showing the bigger values, this time for the maximum ones, is the fifth. The second group also shows higher times than the rest of the groups but in a different scale than the fifth.

- Minimum Equivalent Time of the Blackout Taking into Account the Amount of Service during 12 Months:

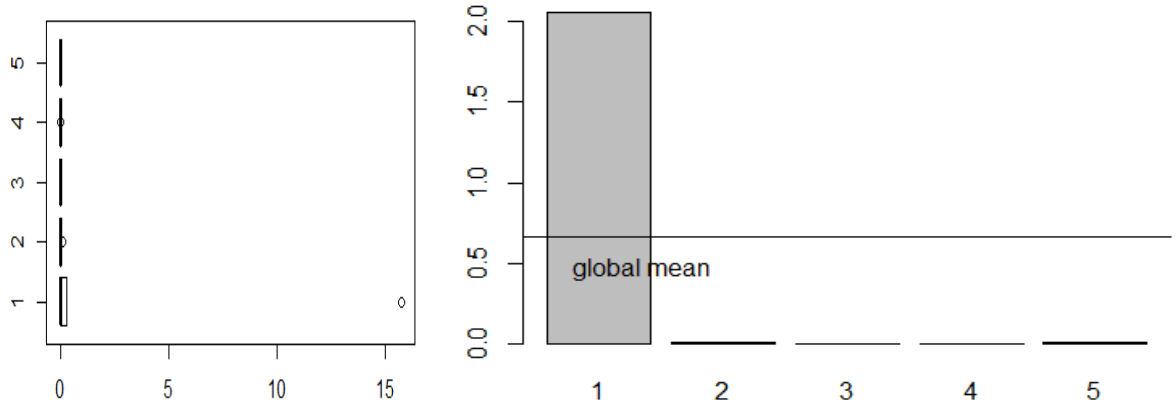


Figure 172

All the groups shows minimum values around zero, even the first one. The reason while it looks very different and bigger from the others is the tiny scale and the outlier present in it. The outliers is the consequence of one longer failure happened in Ireland in 2010.

Climate Variables:

- Type of Climate:

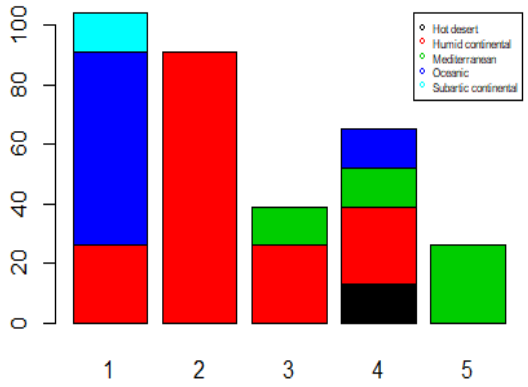


Figure 173

The climate might be a discriminant variable for creating the cluster. As we can see, in every group there is a predominant type of climate, with the exception of the fourth group, which is, very heterogeneous in this sense. The first group is mainly composed by countries with Oceanic climate; the second one only comprises countries with a Humid Continental climate, which is also the predominant climate in the third group. The fifth group is formed by countries with Mediterranean climate.

- Island:

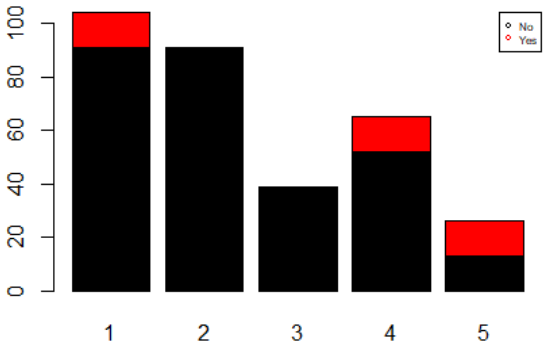


Figure 174

We could expected that the variable Island would not provide with discriminant information when making the clusters. Just three countries are islands in our dataset, namely, Ireland, Great Britain and Cyprus.

- Average Temperature:

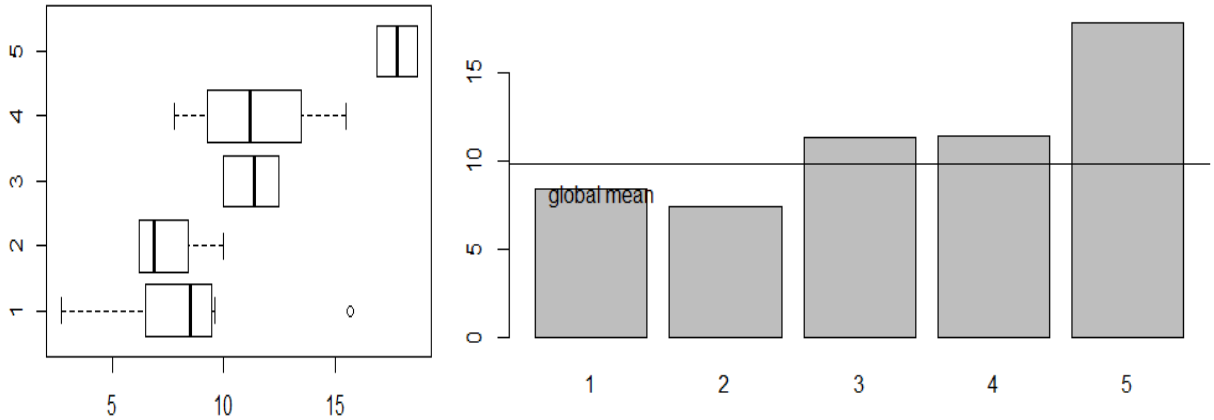


Figure 175

The average temperature of the countries shows big differences among groups and looks like a important variable when deciding the division of the countries. Colder countries are in the second and first group; the hottest ones are in the fifth group and groups three and four have moderated temperatures. The variability in the groups is not very high, showing homogeneity regarding to this variable.

- Average Precipitations:

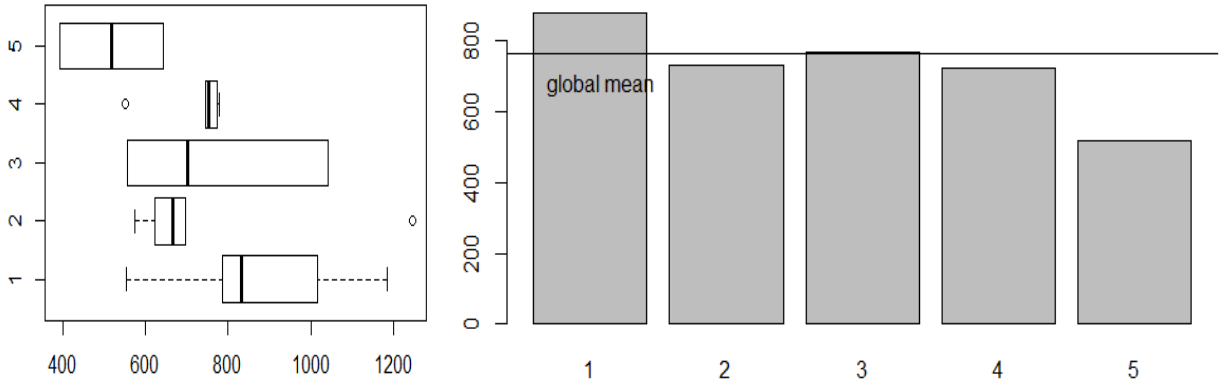


Figure 176

As well as occurs with the average temperature, the average precipitation demonstrate unlike features for every group, even though the variability among countries in the same group is bigger for the precipitation than for the temperature. The rainiest group is the first one, closely followed by the third group. The second and fourth group have moderate precipitations and are more homogeneous as groups. The group number 5 is the driest.

8.4 Appendix d. Principal Component Analysis for the *MEDB* including all quantitative and qualitative variables

Next figure and table shows the factorial map performed for all the quantitative and qualitative variables in the MEDB. Axis represented here correspond to the first and second PC. Some variables (Reason and RatParis) were excluded from the representation before, with the aim to clarify the visualization, as their contribution to the analysis was not decisive.

Colours in the table correspond to the colour of the variables in the map. The abbreviation of the different levels of each qualitative variable used in the representation is also added in the table.

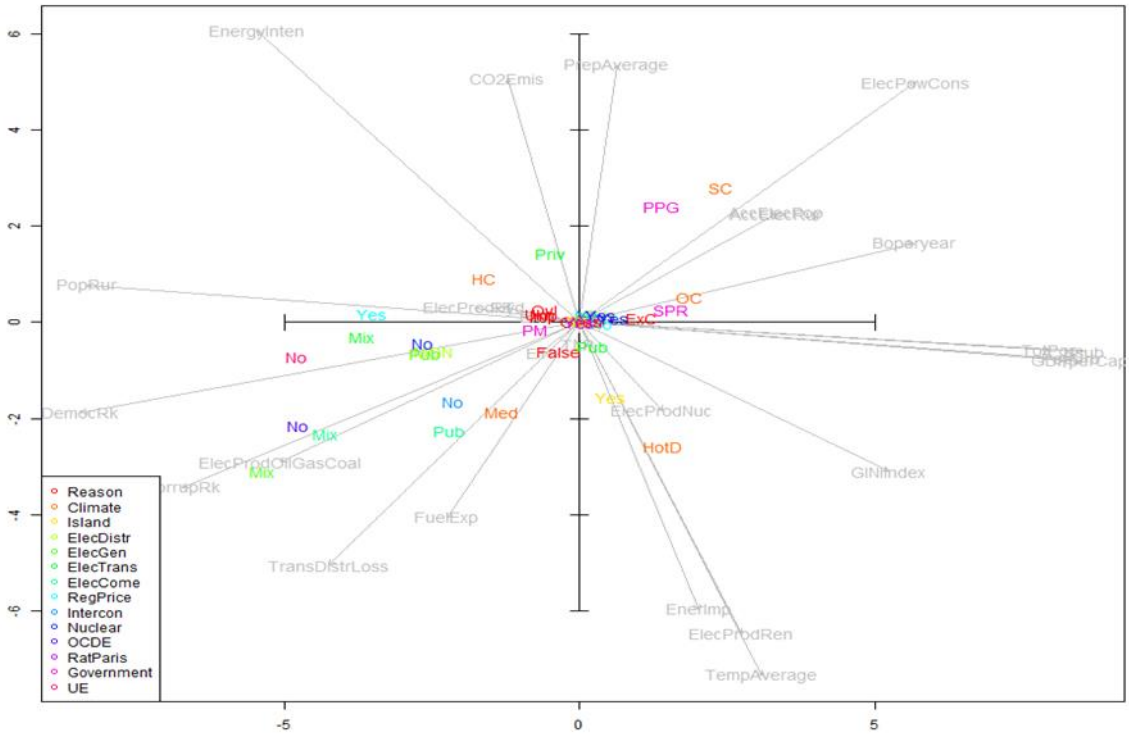


Figure 177

Climate				
Humid continental (HC)	Hot Desert (HD)	Subartic continental (SC)	Oceanic (OC)	Mediterranean (Med)
Island				
Yes		No		
ElectDistr				
Mixed (Mix)	Public (Pub)		Private (Priv)	
ElectGen				
Mixed (Mix)	Public (Pub)		Private (Priv)	
ElectTrans				
Mixed (Mix)	Public (Pub)		Private (Priv)	
ElectCome				
Mixed (Mix)	Public (Pub)		Private (Priv)	
RegPrice				
Yes		No		
Intercon				
Yes		No		
Nuclear				
Yes		No		
OCDE				
Yes		No		
Government				
Parlamentary Monarchie (PM)	Parlamentary Republic (PR)	Presidential with Parlamentary Government (PPG)	Semi-presidential Republic (SPR)	
UE				
Yes		No		
RatParis				
Yes		No		
Reason				
Failure (Fail)	Overload (Ovl)	External Conditions (ExCon)	False Operation (False)	Outside Impacts (Imp)
Unknown (Ukn)				

Table 10